DOCUMENT RESUME

ED 148 809 95 SP 012 081

TITLE Current Directions in Research on Teaching. Conference Series No. 1.

INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.

SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.

PUB DATE Sep 77

CONTRACT NIE-C-400-76-0073

NOTE 88p.; Papers compiled from discussions at a Meeting of the Invisible College of Researchers on Teaching (East Lansing, Michigan, November 17-19, 1976)

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.

DESCRIPTORS *Concept Formation; *Educational Research; *Effective Teaching; *Fundamental Concepts; *Research Methodology; Research Skills; Research Utilization; Time Factors (Learning)

IDENTIFIERS Contextual Analysis; Outcomes of Education; Research Results; *Research Trends

ABSTRACT

This document contains four composite papers, developed from notes and tape recordings, and a collaborative writing effort, on the topics of "time," "context variables," "accumulating findings," and "outcome measures," as they apply to research on teaching. The discussions represented by these papers took place at a three-day conference of sixty-five leading scholars in research on teaching, gathered in an "invisible college." Discussions on "time" investigated conceptions of time in research on teaching, research-based facts about time, hypotheses about time, immediate implications for the practice of teaching, problems (philosophical, substantive, and methodological), and proposed solutions. "Context variables" were investigated from the view of the limits of content, and the concept of "context" from various perspectives (psychological, social psychological, and ecological). "Accumulating findings" sessions discussed problems in accumulating evidence, strategies for accumulating evidence, units of accumulation, issues in weighing evidence, and ways to plan for accumulating evidence. The sessions on "outcome measures" discussed an overview of current research on teacher effects and the problems involved in choosing and assessing outcomes as well as measuring change. Each of the sections contains a bibliography pertaining to the concept under discussion. A list of twenty-nine papers distributed among participants is presented. Appended are an evaluation of the meeting, a list of participants, and a conference schedule. (MJB)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED148809

SP012 081

# Institute for Research on Teaching

College of Education — Michigan State University

Conference Series No. 1

# CURRENT DIRECTIONS IN RESEARCH ON TEACHING

A Meeting of the Invisible College
of Researchers on Teaching
November 17-19, 1976

Published by

The Institute for Research on Teaching
College of Education
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

September, 1977

The Institute for Research on Teaching is funded primarily by the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-76-0073)

CONTENTS


INTRODUCTION . . . . . 1

TIME

by Tracy Harris and Robert Yinger*

Introduction . . . . . 7

Conceptions of Time in Research on Teaching . . . . . 7

Research-Based Facts about Time . . . . . 9

Hypotheses about Time . . . . . 13

Immediate Implications for the Practice of Teaching . . . . . 16

Problems and Proposed Solutions . . . . . 16

a. Philosophical Issues . . . . . 17
b. Substantive Issues . . . . . 18
c. Methodological Concerns . . . . . 19
d. Suggestions for Future Research . . . . . 20

Bibliography . . . . . 23

CONTEXT VARIABLES

by Janis Elmore and Joseph Passalacqua*

Introduction . . . . . 25

The Limits of Content . . . . . 26

Perspectives on Context . . . . . 28

a. Psychological View of Context . . . . . 28
b. Social Psychological View of Context . . . . . 29
c. Ecological View of Context . . . . . 31

Summary . . . . . 33

Bibliography . . . . . 35

ACCUMULATING FINDINGS

by Barbara Perles and Mary Rohrkemper*

Introduction . . . . . 36

Problems in Accumulating Evidence . . . . . 37

Strategies of Accumulating Evidence . . . . . 40

Unit of Accumulation . . . . . 44

Issues in Weighing Evidence . . . . . 47

Ways to Plan for Accumulating Evidence . . . . . 52

Bibliography . . . . . 56

OUTCOME MEASURES

by Steve Olejnik and Susan Wildfong*

Introduction and Rationale . . . . . 57

Overview of Current Research on Teacher Effects . . . . . 57

Problems and Issues . . . . . 59

a. Choosing Outcomes . . . . . 59
b. Assessing Outcomes . . . . . 62
c. Measuring Change . . . . . 65

Summary . . . . . 69

Bibliography . . . . . 71

PAPERS AND REPORTS DISTRIBUTED AT MEETINGS . . . . . 72

APPENDIXES

A. Meeting Evaluation and Analysis--Kent Viehoever . . . . . 75
B. List of Participants . . . . . 78
C. Conference Schedule . . . . . 82

---

*Authors are research interns with the Institute for Research on Teaching.

# INTRODUCTION

Studying the phenomenon of teaching is the mission that has been undertaken by the Institute for Research on Teaching (IRT) at Michigan State University. Founded in April of 1976 by the National Institute of Education, the IRT functions in three capacities: as a center for research on teaching, with studies emphasizing teaching as clinical information processing; as a forum for communication among researchers, practitioners, teacher educators, and educational policy makers; and as a training center for future researchers. Functioning as a forum for communication the IRT sponsors many special conferences and meetings; the meeting of the Invisible College November 17-19, 1976 was one such example.

Research on teaching is a rapidly expanding field within educational research. To pursue this research in a systematic fashion -- avoiding overlap and repetition and keeping abreast of current studies -- it is necessary for researchers to establish a network of communication. Such a network has been formally established at the IRT as an invisible college.

The term "invisible college" was originally used in seventeenth century England to describe the pioneer members of the group which later became the Royal Society of London. Contemporary historian and sociologist of science Derek de Solla Price reintroduced the term to characterize informal clusters of scientists collaborating at newly developed research frontiers. These scientists are geographically dispersed but communicate with each other via formal and informal networks of correspondence, publications, and occasional meetings.

The IRT Invisible College is a group of leading scholars in research

on teaching from universities and research centers all over the United States. The purpose of the college as it has been established by the IRT is to facilitate communication among these scholars, encourage theory-building and cumulative or complementary research in the field, and provide regular opportunities for collection, synthesis, and analysis of research on teaching.

The Invisible College of Researchers on Teaching convened for the first time at Michigan State University, November 17-19, 1976. The approximately 65 college members met in four different sessions and discussed topics of Time, Context Variables, Accumulating Findings, and Outcome Measures as they relate to research on teaching. The following paragraphs describe the nature of each topic/session and indicate the types of issues raised and discussed.

I. Time

Researchers have recently become very interested in time as a variable in studies of teaching and learning. Indeed, time is the central feature of a number of different variables: time on task, academic learning time, length of school day and year, teacher planning of time allocation, and homework time, to name but a few. Time-related variables can be treated as independent variables, dependent variables, and moderator variables. Time can be precisely and easily measured, and time engaged in learning seems intuitively "closer" to student learning than many of the promising variables of the past -- indeed it may be a "missing link" that relates teacher behavior to student achievement.

How do you think time might best be taken into account in studies in teaching? Do you think time has a future? What kinds of activities and

outcomes of schooling might get squeezed out of the school day if time on academic tasks is emphasized more? Do you have data (analyzed or unanalyzed) that include time as a variable?

II. Context Variables

The contexts in which schooling takes place have come to be appreciated both as potentially powerful influences on teaching and learning and as limits for the generalizability of findings in research on teaching. Context variables have been cast as psychological (e.g., student ability), sociological (e.g., group cohesiveness), economic (e.g., SES), organizational (e.g., team teaching vs. self-contained), administrative (e.g., centralized vs. antonomous), political (e.g., democratic vs. authoritarian), architectural, and perceptual. The concept of a "learning environment" within which teaching and learning (and other things) take place needs to be dealt with at a number of levels if research on teaching is to become a more applied pursuit than it is at present. The slippery question of "appropriateness" of using various teaching skills and strategies may hinge on a fuller understanding of the effects of and constraints imposed by context variables. If situations differ as much or more than do persons, we need to learn a great deal more about situations (contexts) in order to make sense of what we know about persons (teachers and learners). Perhaps a useful way to organize our discussion of context variables is to separately consider: a) pupil types, b) subject matter differences, and c) the milieu or learning environment.

What do your data and your experiences tell you about the role of context variables in research on teaching? How can the Follow Through Planned Variations data be brought to bear on this question? Is one of

Cronbach's messages in his paper "Beyond the Two Disciplines of Scientific Psychology" that we have tried to "control" for context variables in a way that makes little sense? How have you measured classroom climate, building climate, etc.? How does the more recently tried ethnographic approach to the study of teaching open up new ways of capturing context variables? Do you see a Context x Treatment Interaction (CTI) interest group emerging in research on teaching?

III. Accumulating Findings

The task of combining the findings of many independent studies in research on teaching has been a reviewer's nightmare. Language differences, definitional problems, methodological differences, and sparse descriptions of treatments, contexts, and procedures have made it exceedingly difficult to cumulate results in such a way that the whole is greater than the sum of the parts. No widely accepted conceptual framework exists within which the field can plan, organize, execute, and report its work. Indeed, research on teaching may not even qualify as a "field" for lack of such a conceptual framework. And certainly, the felt need for theory cannot be satisfied until we have a language within which we can speak and think theoretically.

What might a conceptual framework for relating studies of teaching look like? Should "standards" for executing and reporting studies of teaching be published to increase commensurability, or would such standards reify a paradigm in such a way that scientific creativity might be stifled? To what extent is research on teaching atheoretical? What impact have the several models of the teaching-learning process had on research on teaching? Must literature reviews be limited to studies within a given paradigm?

IV. Outcome Measures

Ultimately, every definition of effective teaching depends on stipulating the "effects" that are to be attributed to the act of teaching. Yet even experienced and thoughtful researchers on teaching effectiveness have invested hours, days, months, and even years in designing powerful treatments (teaching skills and strategies, teacher training) with perhaps insufficient effort to the identification or construction of student outcome measures that are sensitive to good instruction. Measures of the affective outcomes of education are particularly problematic. And, too, better and more fully understood methods for measuring change in studies of natural variation are needed for all types of outcome dimensions.

What can be done to improve the outcome measures situation in research on teaching? Have you constructed any instruments that are especially congruent with particular experimental treatments? What valid uses can be made of standardized achievement tests in research on teaching Do you have any instruments or ideas about outcome measures in the affective domain? What methods of measuring change do you find the most convincing and why?

No papers were formally presented during the meetings, and discussion was loosely structured and largely spontaneous. It was not an easy task to capture on paper and communicate to others the ideas that flowed so freely. Eight IRT research interns took extensive notes at each session, listened to tape recordings of the sessions, then collaborated in writing discussion summaries. (Audio tapes of each session are on file in the IRT Informational Resources Center.) The summary papers, each written by two of the interns and reviewed by the chairperson of the corresponding

session, comprise a synthesis of information and ideas exchanged at the meetings. For obvious editorial reasons and purposes of clarity the discussions and ideas are not always presented chronologically, and some statements made by participants have been elaborated upon and referenced. Nevertheless, the following papers are offered as representative of the discussion and character of the first meeting of the Invisible College of Researchers on Teaching.

# TIME

by Tracy Harris and Robert Yinger

## Introduction

The deliberations of the Invisible College of Researchers on Teaching on the subject of time as a variable in research on teaching left us with four general impressions. First, time is an important and promising variable. Second, time is a very complex variable; there are multiple conceptions of time, and time is seen as interacting with student characteristics, contextual variables, and subject matter or curriculum. Third, the interest in time as a variable in research on teaching indicates that researchers on teaching are beginning to study other aspects of the learning process that influence the behavior of teachers. Finally, researchers have only recently begun to think about, study, and understand the many ways in which time can be conceptualized, operationalized, and used to understand teaching and learning and to improve educational practice.

## Conceptions Of Time In Research On Teaching

One of the most conspicuous aspects of the deliberations on the subject of time as a variable in research on teaching was the variety of ways in which time-based variables were conceived. Time viewed as an independent variable was the most frequent conceptualization. With time viewed as an independent variable, conceptions such as active learning time, time on task, opportunity to learn, and engaged time have been measured or manipulated and related to student achievement.

The second most frequent sense in which time is conceptualized in research on teaching is as a variable to be controlled. This situation is particularly conspicuous in studies of teaching effectiveness in which

two or more experimental treatments or teaching methods are compared. An important concern of the experimenter is that the treatments differ only in the ways intended. If one treatment takes considerably more time to implement than does another, the comparability of effects attributed to the two treatments is threatened. In such cases time is more of a nuisance variable to be compensated for than a factor of central interest in the study.

Differences in teachers' use of time must be contended with in studies of natural variation in teaching. One way to deal with this variability, for example, is to adjust student achievement or some other outcome measure to allow for variations in opportunity to learn, emphasis, or time allocated. Another way to deal with this situation is to characterize teachers according to their pattern or style of time allocation.

The most recent way in which time has been conceptualized in research on teaching is as a dependent measure. Some researchers suggest that time is a useful, measurable, and sensible proxy variable for student learning. That is, although time-based variables are not learning per se, they are so highly correlated with student learning that effective and less effective teaching can be discriminated by measuring these time-related variables.

Another related school of thought contends that active learning time or engaged time is more than a proxy for learning. According to this point of view, student behavior that is classified as "engaged" is the visible evidence of learning taking place. That is, although we cannot see the psychological processes of learning directly, we can see some physical manifestations of it. Therefore, time on task and engagement can be justified as criteria for effective teaching.

The different ways in which time has been used as a variable in research on teaching are illustrated by the brief definitions given below. Each of these time variables could be used as either an independent variable, a dependent variable, or as a variable to be controlled. Two basic ways in which time has been used as a variable in classroom research seem to emerge. The first deals with the teacher's initial control of time and the subsequent apportionment of time to various subject matters, learning activities, and pupils. The terms used to designate this apportionment are:

Time allowed for learning - Carroll (1963)

Opportunity for learning - Carroll (1963)

Time allocated to learning - Wiley and Harnischfeger (1974)

The second basic use of time as a variable focuses on the time that a student spends *actively* involved with a learning task, once time has been set aside for this activity by the teacher. Terms used to describe this process are:

Active learning time - Carroll (1963)

Time on task - Bloom (1973)

Engaged time - McDonald (1976)

Academic learning time - Berliner, Fisher, et al. (1976)

Recently, Berliner et al. (1976) have attempted to refine the notion of engaged learning time by proposing a metric called Academic Learning Time (ALT). ALT is defined as the time spent by a learner engaged in a task that is within an intermediate range of difficulty and is logically related in a direct manner to some specified academic outcome.

Research-Based Facts About Time

Generally, the Invisible College's discussion of time as a variable in

research on teaching evolved from four research-based statements regarding time and school achievement:

1. The quantity of schooling (in terms of days/year and hours/day) is significantly related to student learning and achievement.
2. There is a wide variation in the quantity of schooling available across states, districts, and schools.
3. There is wide variation in the time allocated within classrooms to various subject matters and learning activities, and these allocations are related to student achievement.
4. There is wide variation in engaged or active learning time across students, and there is evidence to suggest that the amount of engaged time is related to student achievement.

Recent work by Wiley and Harnischfeger (Wiley, 1973; Wiley and Harnischfeger, 1974) points to the important relationship between quantity of schooling and student achievement. In a study based on the Detroit Metropolitan Area sixth grade sample of the Equal Education Opportunity Study (EEOS) Wiley and Harnischfeger concluded that in schools in which students receive 24% more schooling than average, the average gain in reading comprehension will increase by two-thirds, and gains in mathematics and verbal skills will increase by more than one-third. Though the specific model that they used has been challenged (see Kurweit, 1976), their results point to the importance of examining quantity of schooling as a variable with significant implications for research and policy.

Other studies have supported Wiley and Harnischfeger's findings. Hyman et al. (1975) showed that the number of years in school is a powerful variable even when SES is controlled. Schmidt (1976), in a recent analysis of the National High School Longitudinal Data, found that even

such a gross measure of time as the number of courses taken in specific subjects is related to achievement outcomes in those subjects.

These findings relating quantity of schooling to achievement become particularly important, and distressing, in light of recent discoveries of large variations of in-school time across states, districts, and schools. Wiley (1973) found that in 1969-70 the average length of the school year in the United States ranged from 172.6 to 182.7 days -- a 5.9% difference. Variation of length of school day may be even greater. Jencks's (1972) reanalysis of the EEOS reveals both mean and standard deviations for length of school year and length of school day for 684 northern elementary schools. (He found much wider variation in length of school day than length of school year.) While almost none of the schools had school years 6% longer than the mean, nearly 16% had school days 10% longer than the mean length. These results are supported by Stallings and Kaskowitz's (1974) finding of variations of up to two hours in the length of school day among 120 schools.

Wide variations have also been reported for the allocation of time within the classroom. In a recent study of six elementary classrooms over a 40-day period Berliner and his colleagues (Fisher 1976a, 1976b) noted large variations in amounts of time allocated to subjects and learning activities within a subject. They found that some teachers allocate twice as much time to a given subject than other teachers and up to five times more time to certain learning activities. These allocations were found to be quite stable over the 40-day periods and seemed to be dictated by factors such as curriculum materials and teacher preferences. (In 1976 Tikunoff and Ward suggested that restrictions on the teacher's time allocations such as school-wide testing and reassignment of students may be so great as to limit the effective school year to as little as 2.5

months of teaching!) It was also found that the amount of time allocated is curvilinearly related to student outcomes, and if academic status was controlled, time allocation factors accounted for up to 10% of the variance in student achievement.

The potential for variance in students' time on task seems much greater than the variance of teachers' allocations. McDonald (1976) reports that the median engaged time for second and fifth grade students in California classrooms was usually under 90 hours for the entire school year in both reading and mathematics. Berliner et al. found that elementary school students were ordinarily engaged approximately 50% of the time. Ward and Tikunoff found no significant differences between percent of engaged time for high and low achievers, though there was a significant difference in the kind of activities in which they were engaged. Other factors affecting engagement that were discussed include setting (Rosenshine) and student preferences (Berliner). There are still questions as to optimal or even realistic engagement times for pupils, though Stodolsky reported that preschoolers have been observed to be engaged for up to 60 minutes at one time.

Data relating time on task to student outcomes is fairly sparse, though studies in progress by Berliner and his colleagues and Harnischfeger and Wiley focus in this area. Stallings and Kaskowitz (1974) found high correlations between time on task and math outcomes with third grade low achievers in the Follow Through Sample. Effects of allocated and engaged time on achievement are also suggested in Bennet's (1976) study of formal and informal classrooms in England.

The data reported thus far on the effects of allocated and engaged time in the classroom are at present merely suggestive. There is a pressing need to examine both the effects of the quantity of instruction and the

mechanisms of instructional time allocation and student engagement. As time is studied more closely its influence and effects appear increasingly complex. Researching the role that time plays in teaching and learning will not be a panacea for understanding classroom processes, but it should provide at least some understanding of an influential mechanism affecting classroom life.

Hypotheses About Time

The notion that time makes a difference was the overriding hypothesis implicit in the discussion of time as a variable. Also, time is a variable in the sense that it is a metric for describing the classroom, and so the second major hypothesis is that time is a useful metric for describing the classroom. From these two hypotheses stem many questions and considerations about time as it relates to teacher decisions, teacher and student behavior, achievement, the process of learning, the nature of the task, the nature of the student, and the context of learning and teaching.

Just as conceptions of time can be categorized as independent variables, dependent variables, or as variables to be controlled, so can the research questions on time be categorized. The deliberations on time as an independent variable stemmed from the hypotheses that variations in time allocation produce variations in student outcomes. If time is a resource to be divided among tasks, then allocating time to one task limits the time available for other tasks.

Several factors might be involved in determining the amount of time allocated per task. The teacher may be directly instrumental in time allocation decisions based on his or her beliefs about the desirable outcomes of schooling and subsequent curriculum choices, etc. Preliminary

findings from ongoing research (Berliner et al., 1976) suggest that there is more within-class variation in the amount of engaged time in math than in reading. Perhaps this can be attributed to the homogeneity of reading groups as compared to the heterogeneity in math from lack of grouping. These allocation decisions may be determined largely by factors outside direct teacher control such as class size, student traits and consequent needs, state regulations, and length of school day. Major research questions deal with the relationships between these areas. For example: are the effects of a shortened school day more pronounced on elementary than high school students? Is there consistency among teachers in the amount of time allocated to different content areas? Do time allocations contribute to the effects of curriculum materials? These are a few of the questions to be answered.

The concept of time as engaged time is hypothesized to be related to achievement outcomes. As a function of allocated time (see note below)* engaged time is subject to similar considerations such as teacher decisions, content areas, context issues, and student traits. Example research questions are: What is the relationship of engaged time to the amount of content covered? How does the redundancy of materials affect student engagement? Can we assume that the teacher's task is to help students stay on task? It is postulated that the nature and setting of the task determine what the child can and will do and that teacher instructions affect time on task. A particularly intriguing hypothesis is that as the number of disengaged students increases a point is reached where all other students must become disengaged.

---

*Conceivably, there are students who generate their own variety and learn to ignore the teacher. This adds another variable to the equation of engaged time being a function of allocated time.

Time as a dependent variable can be considered a proxy for learning. ALT may be a process measure of learning. This use of time as an outcome measure suggests the important concern of determining the optimal amount of engaged time for learning. Berliner and his colleagues (1976) have collected data that indicates that many second graders may, on an average, be spending less than 130 hours per year in reading and math combined. There are also indications that some individuals may be spending less than 40 hours per year in any formal or direct learning tasks. There is some concern expressed that the time spent on task is not sufficient to achieve the kinds of gains desired by the general public. Some claims are made that there are many students in classrooms who do not know what to do and so their time is not being spent learning. Perhaps ALT would be a good measure for giving teachers feedback on their instruction, allowing them to determine if enough time is allocated for certain tasks and whether that time is being used efficiently by students.

The relationship between time and learning is suggested by the use of engaged time as a measure for learning. The more time spent learning, the more one learns. Little time spent on task suggests either less learning or quicker learning. Researchers might also investigate allocated and engaged time spent at home to understand better how time relates to learning. Perhaps a student spends little time on task at school and yet learns because he studies at home, not because he learns quickly as one might determine from classroom observation. It is possible that transfer effects mediate the relationship between allocated and engaged time and increased learning so that time spent on one task decreases the amount of time necessary on another task.

In summary, time as a measure of aspects of the classroom generates

many useful research questions or hypotheses. Time has to be contended with as both an independent and dependent variable. It is a concept which will reduce to a few effects, giving researchers and practitioners better control on the processes and products of the classroom.

Immediate Implications For The Practice Of Teaching

One consideration in any study of teachers is what use the questions and answers generated by research will be to the practice of teaching. Time as a variable in research has several implications for teachers. As a metric for the events of the classroom it can be used by teachers to maintain or establish awareness of the processes and products of school. Creating an awareness of the effects of differential time allocation should aid the practitioner in making decisions. Preactive decisions may be made as to curriculum and materials and content and setting (etc.). Preactive decisions can be made based on minute-by-minute or day-by-day feedback from the class. Time as an independent variable may be manipulated to advantage. The teacher can use ALT, for example, as a personal criteria for determining the impact of specific teaching methods and materials. Time as a dependent variable allows the teacher to better assess outcomes and processes of the classroom. Also, time is a factor readily considered by practitioners, and this increases its worth as a variable for study.

Problems And Proposed Solutions

The use of time as a research variable may alleviate many problems inherent in research on teaching, but it also creates its own. There are philosophical issues and values which influence what is studied as well as the conclusions drawn. There are substantive issues concerning what is

really being measured by time, and there are methodological concerns about how one uses time as a measure.

Philosophical issues. The one overriding philosophical consideration is the question of what a student (or a teacher or school) *should* do. Most of our research is naturalistic and so we discover what is being done in the classroom. But the issue of values must be confronted. We must look at the purpose of education and how our beliefs are bound by our ideologies. It does not suit our purpose here to address the broad issues, but rather to consider several specific, practical aspects of value-based decisions about time.

It is generally believed that a teacher's goals affect time allocations. For example, one teacher may stress reading over math; another may stress decoding over comprehension in reading. One teacher might use recess as a reward for work well done (or, at least, work done) while another uses recess as an integral part of the learning experience, i.e., everyone goes to recess regardless of work completed. Often one's beliefs are based on tradition or myths of how time is to be used. One such belief is that it is the teacher's job to keep the student task-oriented. By adult standards not much time in the elementary classroom is spent on task. Is this detrimental to our goals?

Some teachers have the goal of creating or maintaining intrinsic motivation, thus making themselves dispensable. Others see themselves as facilitators and resources for learning.

Teachers' beliefs influence the life of the classroom and the domains of research. There are those who insist that research can be value-free and that philosophical issues are for philosophers and perhaps practitioners. But we must consider how our beliefs affect what we study and how we study it. The philosophical issues lead to substantive ones. How are we to examine time to answer to our values?

Substantive issues. Mentioned above was the observation that not much time in school is engaged time. This leads to the question of whether time *not* spent on task is wasted time. When research measures Academic Learning Time, for example, is time off task not measured because it is seen as unimportant, or because it is not easily categorized by its object? In other words, is time off task not measured because it is difficult to determine what a child is doing when he's not engaged in a task? Not only is there a question about the amount of engaged time expected of a student, but also a question of how much engaged time a student can handle.

An issue stemming from values as they related to time allocation differentially produces outcomes. If more time is spent in reading, then less time is allocated for math, and consequently, less is learned about math. This means that optimization of time for all subjects is not possible. This problem is complicated by a possibility that when engaged time is treated as an outcome the less time spent the better. Fast learners will need less time, which makes it necessary to consider the interactions between engaged time, outcome, and learner ability. Another complication is the process of transfer which would suggest that the more time spent now, the less time needed later.

Complications arise when examining the relationship between allocated time and engaged time. Engaged time is not completely determined by each individual student or by the allocated time. There may be a time when those who want to be engaged cannot be. It is likely, for example, that when the majority of students in a class are not engaged, those who are engaged will be interrupted by these students or by the teacher's management techniques, such as having everyone put his or her head down.

The sequence in which subjects are taught may be a factor in the

amount of time allocated, needed, or engaged. If a task is allocated for the end of the day it may never be reached because of unforeseen circumstances and time considerations. If a story about a pet is read early in the day it might facilitate writing stories or drawing pictures of pets later in the day.

We have been looking at time, both allocated and engaged, as it relates to and affects the students' behavior. How do teachers spend their time when they are not directly instructing? What about teachers' allocation of their own time and their consequent engagement? Not only do students study outside of the classroom, but teachers as well. We cannot confuse teacher behavior as it relates to outcomes with time allocation (for students) as it relates to outcomes.

We have come to realize that time can be used to measure many aspects of the classroom and that time as a variable in research becomes more complex as it is examined closer. Now we shall move to consider ways in which one measures and analyzes these aspects through time.

Methodological concerns. Several methodological issues are of considerable importance for the study of teaching in general, and more specifically, for time as a research variable. Because they are issues which we all have encountered before this discussion will be little more than a mention of each as it relates to time as a metric based on completed research.

The validity of observational methods in the lower elementary grades (K, 1, and 2) presents few problems. Children in these grades apparently have no guile. Observers have little trouble determining whether or not a second grader is engaged in a task (Berliner et al., 1976). Older children, however, do present problems, and it is harder to be certain of

the validity of observations. The training of observers affects reliability, as always.

Another methodological issue is the use of volunteer teachers. Dependent on the hypotheses of a study, this need not be a problem in determining allocated time in that it gives a liberal test. In Berliner's et al. study the times were probably overestimated, yet they were still lower than expected.

Teacher logs may be used to get a better sense of what teachers do, but this method is apt to influence the subsequent time allocations by making teachers aware of their behavior. Teachers must be trained to use logs for the most benefit to the researcher.

Standardized tests are limited as outcome measures because teachers might stress skills other than those the tests measure.

Of course, sampling and the generalizability of findings are both considerations for studies of teaching. Also, the content area must be considered in any data analysis as well as the setting for teaching. Another analysis consideration is the use of mean estimates as compared to analyzing individual differences. The above-mentioned methodological questions and issues have certainly not been resolved yet, but the fact that they are considered in the research is a step forward.

Suggestions for future research. Researchers are again using time as a metric for studying the classroom. There is still much to be discovered about time which should prove useful for understanding and prescription of teaching. Ideas are developing which focus on the future possibilities inherent in the consideration of time. The following paragraphs describe several issues requiring further study:

Just as there is a limit to the amount of time which can be allocated

to any task because of the length of the school day and number of subjects to be covered, there is a theoretical limit to the amount of time during which a student can be engaged in a task. These limits must be explored before policies are based upon them. These limits are mediated by factors such as student motivation, quality of instruction, and the developmental age of the student. The ratio of engaged time to allocated time needs to be explored to ensure the optimum amount of time allocated to each task.

Time as a metric is value-free. It is a measure which means any value it has is determined by what it measures. A case in point is the measure of time on task. Not all tasks are worthwhile. Attention must be paid to quality and value issues when doing research with time as a variable. If the task is worthless no amount of time spent on it will benefit the student. No time, then, is the goal.

Ideally, the value of our research on teaching with time as a specific focus is determined by its usefulness for intervention and prescription. Policies must be based on knowledge about what children and teachers can do and on social factors such as the community's requirements and the requirements of the school as an institution. One idea for future research, then, is to go directly from the pilot study to determining validity rather than proving validity before implementation. Rough notions of the possible effects are sufficient for moving on to the investigation of effects through intervention. Without this leap from pilot to intervention too much time is wasted when action is needed.

Engaged time as a proxy for learning is considered a promising variable. Its measures, though, are still in need of refinement. For

example, addressed earlier was the point that not all engaged time is spent at school, but that time spent engaged at home or in the library must certainly mediate the relationship between time on task in the classroom and other process and outcome variables. Allocated time is multi-determined and more study is necessary on its determinants. Transfer studies would allow for more consideration of the relationship of time allocation to outcomes.

In summary, then, future research can fruitfully use time as a measure of the teaching process. Of pressing importance are the issues of the student's ability to be engaged, the content and value of the task, the context of the task, and teacher's role in the classroom. One role of research is to develop policy, and so intervention must be considered as an immediate consequence of research findings.

Bibliography

Bennett, N. Teaching styles and pupil progress. London: Open Books, 1976.

Berliner, D.C., et al. Proposal for Phase III-B of the BTES. San Francisco: Far West Laboratory for Educational Research and Development, 1976.

Block, J. Time in school learning: An instructional psychologist's perspective. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., 1975.

Bloom, B.S. Time and learning. Thorndike. Address, 81st Annual Convention of the American Psychological Association, Montreal, 1973.

Bond, G.L.,& Dykstra, R. The cooperative research program in first grade reading instruction. Reading Research Quarterly, 1967, 2, 5-142.

Carroll, J.B. A model for school learning. Teachers College Record, 1963, 64, 723-733.

Carroll, J.B.,& Spearitt, D. A study of a model of school learning (Monograph No. 4). Center for Research and Development in Educational Differences, Harvard University, 1967. (ERIC Document Reproduction Service No. ED 045 477.)

David, J.L. Summer study: A two part investigation of the impact of exposure to schooling on achievement growth. Unpublished doctoral dissertation, Harvard University, 1974.

Fisher, C.W. A study of instructional time in grade 2 mathematics (Tech. Rep. II-3). San Francisco: Beginning Teacher Evaluation Study (BTES), Far West Laboratory, 1976.

Fisher, C.W. A study of instructional time in grade 2 reading (Tech. Rep. II-4). San Francisco: BTES, Far West Laboratory, 1976.

Harnischfeger, A.,& Wiley, D.E. Teaching-learning processes in elementary school: A synoptic view (Tech. Rep. 75-3-1). San Francisco: BTES, Far West Laboratory, 1975.

Harnischfeger, A.,& Wiley, D.E. Notes for a field study of instructional time. (Tech. Rep. 75-10-11). San Francisco: BTES, Far West Laboratory, 1975.

Harnischfeger, A.,& Wiley, D.E. Conceptual policy issues in elementary school teaching-learning process. Research proposal submitted to N.I.E. Basic Skills Group, 1976.

Harnischfeger, A., Pifer, R.E.,& Wiley, D.E. Instructional time allocation in fifth grade reading. San Francisco: BTES, Far West Laboratory, 1976.

Harris, A.J., Morrison, C., Serwer, B.,& Gold, A.L. A continuation of the CRAFT project: Comparing reading approaches with disadvantaged urban Negro children in primary grades. New York: City University of New York, 1968. (ERIC Document Reproduction Service, No. ED 010 297).

Harris, A.J., & Serwer, B. Comparison of reading approaches in first grade teaching with disadvantaged children. New York: City University of New York, 1966.

Hess, R. & Takanishi, R. The relationship of teacher behavior and school characteristics to student engagement (Tech. Rep. 42). Stanford: Stanford Center for Research and Development in Teaching, 1974.

Husen, T. (Ed.) International study of achievement in mathematics: A comparison of twelve countries (Vol. 2). New York: John Wiley and Sons, 1967.

Hyman, H., Wright, C., & Reed, J. The enduring effects of education. Chicago: University of Chicago Press, 1975.

Instructional time allocation in fifth grade reading. (Tech. Rep. II-5). San Francisco: BTES, Far West Laboratory, 1976.

Jencks, L.S. et al. Inequality: A reassessment of the effect of family and schooling in America. New York: Basic Books, 1972.

Karweit, N. Quantity of schooling: A major educational factor? Educational Researcher, 1976, 5 (2), 15-17.

McDonald, F.J. Beginning Teacher Evaluation Study, Phase II (Summary Report). Princeton, N.J.: Educational Testing Service, 1976.

Rosenshine, B. Classroom instruction. In N.L. Gage (Ed.), The psychology of teaching methods, NSSE Yearbook, 75:1, Chicago: University of Chicago Press, 1976.

Schmidt, W. An unpublished analysis of the National High School Longitudinal Data. Personal Communications, 1976.

Stallings, J.A., & Kaskowitz, D. Follow-through classroom observation evaluation, 1972-1973. Menlo Park, California: Stanford Research Institute, 1974.

Tikunoff, W.J., & Ward, B.A. Some selected findings from three studies. Effective teacher education program. Far West Laboratory, 1976. (mimeograph)

Wiley, D.E., & Harnischfeger, A. Explosion of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. Educational Researcher, 1974, 3, 7-12.

Wiley, D.E. Another hour, another day: Quantity of schooling, a potent path for policy. Studies of educative processes (No. 3). University of Chicago, July, 1973. Also in Sewell, W.H., Hauser, R.M., & Featherman, D.L. (Eds.) Schooling and achievement in American society.

# CONTEXT VARIABLES

by Janis Elmore and Joseph Passalacqua

> *No act can be performed independent of some context. The meaning of any act depends upon how it relates to the context, what consequences it produces. No attitude, philosophy, feeling, belief, or habit pattern occurs de novo. All such states arise from interaction between the person and his environment. No human experience can be divorced from some environmental context. The importance and influence of the environment cannot be exaggerated.* (Carpenter & Finley. The Skinner Primer. London: The Free Press, Collier & Macmillan, 1974.)

## Introduction

The context within which teaching and learning take place has become a major area of concern for researchers on teaching. Participants of the second session of the Invisible College meetings attempted to deal with this concern and its impact upon educational research. It became clear during the deliberation on this subject that contextual factors influence the generalizability of research findings and further set limits on the application of the research. The discussion began with the contextualization of time, but the differing perspectives and conceptions of context forced the deliberators to be concerned with the question, "What is meant by context?" This review will attempt to summarize the varying perspectives of the participants and the current methodological and coding problems that were discussed. Presentation of each perspective includes an analysis of related research hypotheses and concerns. It is important to note that the perspectives are rooted in various fields of inquiry, i.e., psychology, social psychology, and sociology. The representative perspectives are additive, social environmental, and ecological and cover a wide range of variables in relation to contextual issues and research on teaching.

The Limits Of Context

Although each of the perspectives of context mentioned (psychological, social psychological, ecological) offers a different prescription for studying context, all of the views of context must contend with the underlying problem of limits determined by contextual issues. The context of the classroom places limits on the behavior in that classroom (Brophy and Evertson, 1976). Context influences behavior and environment.

The appropriateness of actions or inactions in the learning environment are also dependent on context. For example, the appropriateness of a fight in a classroom is dependent on the context of the classroom. One class may be extremely upset -- fights may be unheard of in the context of the classroom. Yet, in another classroom, fights may be a common occurrence and the class may remain undisturbed. To avoid a misinterpretation of the meaning or importance of an incident, coding schemes have to be developed that reflect the context in which the incident took place. The appropriateness of a fight in a classroom is not dependent on the teacher, but on the context of the classroom (which may or may not be affected by the teacher and his/her reaction to the fight). The possibility that outcome measures may be independent of teacher effects in the classroom, then, is something researchers on teaching must consider.

Related to the problem of attributing teacher effects to outcome measures in the classroom is the problem of attributing outcome measures to the effects of the individuals or the effects of the context of the situation. The relationship between individual behavior and context can be studied, yet it is impossible to separate the two, cleanly dividing the effects attributable to each. The person to a certain extent creates the context (Barker, 1968).

It is this relationship between individual behavior and the context, then, that should be investigated. There is a need to conceptualize context in relation to human behaviors. Rather than dissecting the relationship and studying how much each fiber influences the other, researchers should build a supra structure or network to describe and characterize context.

Efforts to establish this conceptualization of context have already begun. Brophy and Evertson (1976), in an effort to define the limits imposed by context and unify research on teaching, have distinguished several important contexts in the classroom. Their suggested methodology includes combinations of the following:

1. Coding data from different contexts separately (small groups vs. whole class; academic vs. procedural vs. behavioral; various specific contexts in first grade reading and in junior high English and math.

2. Coding teacher-initiated interactions separately from student-initiated interactions.

3. Retaining the sequence of action and reaction in extended interactions.

In addition, Brophy and Evertson (1976) have found grade level to be an important consideration in context. The behavior setting of a first grade class differs markedly from that of a junior high class. The importance of the conceptual differences in pacing and group size and teacher-student interactions are also well discussed in their article.

Although context imposes limits on expectations and perceptions, it also places limits on the generalizability of the findings of research on teaching. Doyle's (in press) example of educating teachers to establish eye contact with individual students illustrates the problems of generalizing research on teaching without accounting for the context. Eye contact is a positive aid in small groups, but in a classroom, establishing eye contact

with one student excludes the rest of the class and may create the potential for disruptive activities or inattention. If this is the case, then prescribing eye contact to teachers without prescribing the context may hinder rather than aid the teacher's effectiveness in the classroom. Although context limits the generalizability of research findings, prescriptions which account for context will be more reliable and relevant to teachers than previous research findings.

Perspectives On Context

The discussion began with contextualizing the area of time but soon digressed to a deliberation over appropriate coding schemes for differing subject matter. Participants noted the value of coding both math and reading separately. As pointed out by one participant, researchers who have this type of data have failed to include it within their analysis. This loss of information further limits the generalizability of findings. Thus, it was suggested that future research maintain separate codes for differing subject matter in the final analysis and report. (This debate pertains to contextual issues if one recognizes that for these researchers subject matter is the context, e.g., the reading group.)

Psychological view of context. The psychological analysis of teacher and student behavior was also an interest expressed by researchers. This type of research analyzes teacher behavior as a function of personal capability and motivation. In addition, the psychological analysis stresses the use of process variables, including many variables often treated as if they were generic teaching variables (Brophy and Good, 1970). Generic variables, according to Gage (1972), are those which have generalized effects independent of specific circumstances. This model conceptualizes variables as if they applied with equal relevance and appropriateness to

all classroom contexts when in fact they do not. (This argument reaches beyond the scope of this review. For further clarification see Brophy and Evertson, 1976.)

The psychological model has failed to take into account the relevant contexts in which learning takes place. One discussant stressed not only the importance of analyzing differing subject matter, but stated that further emphasis must be placed on coding the setting. This entails an analysis of contextual variables which effect behaviors (e.g., age, grade level, quality of subject matter, group size, lesson objectives, and differing achievement levels of students). During the deliberation the participants voiced a concern for coding the appropriate levels of difficulty. This type of coding is congruent with the belief that what is appropriate for a child may vary from setting to setting. Thus, coding schemes that include these various levels of appropriateness should help to increase the validity of research findings in this area.

This approach to coding schemes raises the problem of determining which variables are the appropriate ones to study and what is the appropriate number of variables to examine. Several discussants expressed an interest in studying as many variables as possible. Others were concerned with the most influential variables, i.e. SES, subject matter, time on task. The discussants arrived at no specific answers of these questions and pointed out that there are no studies being conducted that address these issues. A majority of the participants advocated future research that attempts to answer these questions.

Social psychological view of context. Evolving from the psychological view of context the social psychological approach is represented by Walberg's work. Criticizing studies of context that focus on low-inference measures

(process variables) and neglect high-inference measures, Walberg is concerned with studying the affective environment of the class using the child's perception of the situation. He stated during the discussion that the child should be treated as the independent variable and the context should be viewed through the child's perceptions of the situation, implying that if researchers ignore what is going on in a child's mind they may miss the picture or context. The importance of the child's perception is upheld by research-based facts which have shown that perception is highly correlated with achievement.

In the social psychological approach to studying context researchers coding the differing levels of complexity of subject matter would ask the child performing the task his/her perception of the subject matter. Using this method the researcher would be able to explore differing levels of learning, including the learning influenced by the affective environment. Although many current studies center on the child's perception of context, the social psychological view is more generally concerned with studying the norms operating in the social context of the classroom. To understand these norms, high-inference measures of the affective environment and group cohesiveness are stressed. The child's perception of the situation is important to the understanding of the norms of the classroom, but this perception is only one aspect of the social psychological perspective.

Another essential index of the affective environment of the classroom is group cohesiveness. This index is in contrast to the psychological perspective, where cognitive outcomes or achievement alone are often emphasized. One goal of the social psychological view is to optimize group cohesiveness with the teacher functioning as an integral part of the class, sensitive to everything going on in the classroom. One study illustrating

an attempt to optimize group cohesiveness concentrated on supplying the teacher with immediate feedback from the students in the class by using a red light, green light technique. This feedback was intended to aid the teacher in optimizing group cohesiveness by allowing him/her to become more sensitive to the students' needs. This study raises questions such as how many cues can a teacher attend to at once, and how context-dependent is this ability? Current studies have not yet answered these questions.

Thus, the social psychological perspective offers an alternative way to study context based on the study of the norms of the classroom. More specifically, this view stresses the study of the affective environment of the classroom. Two important indexes of the affective environment are the child's perception of the context and the evaluation of group cohesiveness. These high-inference measures are the major emphasis of the social psychological model.

Ecological view of context. Another perspective of context is the ecological view. From this ecological perspective, learning to teach involves learning the texture of the classroom and a set of behaviors congruent with environmental demands (Doyle, in press). Further, this view stresses the theory that environmental demands interact with performance to shape observed behavior and to establish limits on the range of response options (Doyle, in press). This model defines context as the classroom environment and differs from the psychological view by labeling teacher decisions as functions of the environmental demands, not personal capability and motivation. Clearly, this view emphasizes the context within which behavior occurs.

Two future research concerns that this model must analyze are: 1) What are effective educational environments? and 2) What is effective

teacher behavior? These questions must be studied if this model is to move from its infancy -- trying to identify variables that make up the environment -- to a more mature stage where prescriptions for a positive educational environment can be administered. Answers to the questions raised about the ecological model's effectiveness in writing prescriptions that deal with teacher decision making and teacher responsibility depend upon future educational analysis involving this perspective.

Consistent with this ecological view of context and its relationship to the demands of the situation is another perspective -- one which differs in its research emphasis. This perspective is a conceptual attempt to arrive at some genotypical dimensions of classroom activities that make a difference in how children behave and in how teachers might direct activities dependent upon the properties of activities (Kounin and Doyle, 1975). This view defines the context variable as activity or behavior settings (e.g., math and reading group) which have two features: a particular milieu with its appropriate behavior objects and a format or program with a standing behavior pattern which fits that milieu (Kounin and Doule, 1975). The activity setting of a symphony was offered as an example of how environment controls behavior. People who attend a symphony in one part of the country act the same as people in another part of the country; for example, individuals attending a symphony concert in New York will exhibit the same behaviors as those attending a concert in Detroit. The norms operating within this activity setting require the participants to act in a similar manner -- a manner which is appropriate to the symphony.

Ecological models must be considered in which properties of activities are seen as essential parts of school environments and have an impact upon behavior of both pupils and teachers. The behavior setting is the context, which is either a math or reading group. The social climate is not identified

as a context because it is not a behavior setting with its own props and programs, according to Kounin. The context of the school, i.e. social methods of behavior or behavior controlled by setting, are not contexts in this model because they do not meet the definition of behavior settings.

There appears to be a distinct difference between the research emphasis of the two ecological perspectives. The former one is interested in the psychological aspects of context while the latter is concerned with the physical restrictions of the behavior settings. Also, the first one considers the processing of information to accomplish tasks while the other views the active mental life of the student or the teacher in the context of the classroom setting.

Despite these differences both ecological views of context support further research on the norms of the classroom. This research should include the study of teacher techniques that maintain the activity setting in the classroom as well as research on the coercive effects of the behavior setting. With further research and definition the ecological view of context could offer a unique understanding of the dynamics of the classroom.

Summary

One of the major problems in discussing context is the differing criterion attributes. This problem, although shared by most researchers, is not commonly recognized; therefore, discussions of context cannot proceed in any reasonable form, preventing the development of specific criterion definitions. There is an urgent need for a specific level of understanding and agreement on the meaning of context. This definition of context is crucial to further development of research on teaching.

It is obvious that differing methods of inquiry have different meanings for context. From the psychological perspective of context, the factors that moderate teacher-student relationships (e.g., student variables, student socioeconomic status, I.Q., teacher qualifications) are defined as the context. In contrast, the social psychological perspective focuses on studying the norms operating within the social system. The third view discussed -- the ecological view of context -- concentrates on the description of behavior settings and their effects on the actions of people within them. Within this view, the environment is seen as creating physical as well as psychological constraints.

Research on teaching must consider contextual variables. If context is ignored the generalization of research findings may be inappropriate. Practical use of prescriptions written for the phenomena studied will also be inhibited by research which lacks context variables. Context clearly places limits on the behaviors, perceptions, and expectations of those involved in the learning process.

Thus, regardless of the differing meanings and levels for studying context, researchers must recognize context as an important and influential unit of analysis.

Bibliography

Barker, R.G. Sociological psychology. Stanford, California: Stanford University Press, 1968.

Brophy, J.E., & Evertson, C.M. Context effects on classroom process variables (Rep. #76-10). Austin, Texas: University of Texas Research and Development Center for Teacher Education, 1976.

Brophy, J.E., & Good, T.L. Teacher-child dyadic interactions: A new method of classroom observation. Journal of School Psychology, 1970, 8(2), 131-138.

Carpenter, F. Skinner primer: Behind freedom and dignity. New York: Free Press, 1974.

Doyle, W. Summary of learning the classroom environment: An ecological analysis of induction into teaching (DEF). Denton, Texas: North Texas State University, 1976. (mimeographed)

Doyle, W. The uses of nonverbal behaviors: Toward an ecological model of classrooms. Merrill-Palmer Quarterly, in press.

Gage, N.L. A factorially designed experiment on teacher structuring, soliciting, and reacting. Journal of Teacher Education, 1972, 27, 35-38.

Kounin, J.S., & Doyle, P.H. Degree of continuity of a lesson's signal system and the task involvement of children. Journal of Educational Psychology, 1975, 67, 159-164.

Kounin, J.S., & Gump, P.V. Signal systems of lesson settings and the task-related behavior of preschool children. Journal of Educational Psychology, 1974, 66, 554-562.

Rosenshine, B. Current topics in research on teaching. Urbana, Illinois: University of Illinois, 1976 (mimeographed).

# ACCUMULATING FINDINGS

by Barbara Perles and Mary Rohrkemper

## Introduction

The need for systematic accumulation and interpretation of educational research findings is evident. While many studies have been completed, their findings have rarely been accumulated in useful ways. In education, the phenomenon of the one-shot, never-read-or-tried again doctoral dissertation is simply too common. To continue these individual, isolated research studies without either deriving generalizations across the studies or viewing them within a meaningful framework seems fruitless in a field which has both understanding and intervention as its goals. Studies yet to be defined can only broaden and deepen our knowledge if they can build on what has preceded them. Obviously, such a building can not occur without a familiarity with others' research and a healthy sense of "what it all means." As Gene Glass (1976, p. 4) has remarked, "Our problem is to find the knowledge in the information ... the knowledge that lies untapped in completed research studies."

The purpose of the Accumulating Evidence session of the Invisible College meeting was to address these concerns as confronted by those within the field of research on teaching. Throughout the session, it was evident that the participants' perceptions of the overall purposes of accumulation influenced the direction and content of the discussion. Of the many possible uses of accumulation of research findings -- immediate decision making, direction of future research, verification of previous findings, content of teacher training programs, theory building, etc. -- the foremost concern of the participants was the role of research accumulation

for planning the content of teacher education programs. The following summary reflects this perspective and is organized into five topic areas:

1. Problems in accumulating evidence
2. Strategies of accumulating evidence
3. Unit of accumulation
4. Issues in weighing evidence
5. Ways to plan for accumulating evidence

Problems In Accumulating Evidence

Two problems in any accumulation of research are the difficulty in interpreting the researcher's selected variables and the small overlap of these selected variables among different investigators. This inconsistency in variable definition and use creates havoc for the would-be accumulator. Discussion of the difficulties in interpreting and understanding an investigator's variables focused initially on the issue of low-inference vs. high-inference variables. Lively debate over the optimal level of abstraction of variables ensued.

Some participants advocated low-inference variables such as "number of questions asked," and "amount of seatwork time allotted." They asserted that low-inference variables are easily defined and lend themselves to quantification, thus facilitating accumulation of research. If investigators used low-inference variables, they maintained, others would be able to understand their exact meanings.

Other participants expressed discomfort with low-inference variables, stating that these variables sacrifice validity for reliability. It was argued that the justification for low-inference variables (grounded in the assumption that such variables are related to the more abstract behavior

the researcher is studying) is questionable. An experimenter who observes the frequency of a low-inference variable must assume that this readily generated tabulation represents the occurrence of the more general phenomenon. For example, assume an investigator is interested in the impact of teacher warmth on the classroom climate. An experimenter who believes that "warmth" is too abstract a notion for reliable systematic inquiry, and asserts that the phenomenon "warmth" is adequately represented by the occurrence of "smile," may choose to confine his measurement to the frequency of teacher smiling. Thus, the researcher assumes that his data on the frequency of teacher smiles can be understood and interpreted as teacher warmth. This basic assumption of a direct and invariable relationship between such aspects of behavior as "smile" and "warmth" was strongly questioned by some participants. One discussant asserted that such an assumption was incorrect in that it mistook a component behavior (e.g., smile) of a more general phenomenon (e.g., warmth) to be synonymous with that phenomenon. It was argued further that the tabulation of the frequency of occurrence of a low-inference behavior does not necessarily imply the occurrence of the high-inference behavior.

Those participants who advocated the use of high-inference variables in research and accumulation made the following assertions: First, high-inference variables are more systematically related to outcome; second, they are more stable across lessons than are low-inference variables; and third, they are more readily understood by teachers and teacher educators since high-inference is the level of abstraction at which persons normally operate. Concentration on low-inference variables in classroom research efforts, it was maintained, could only occur at the expense of a more complete understanding of classroom phenomena.

Debate turned to the issue of the appropriate inference level of the research information provided for both preservice and inservice teachers. Questions raised included: Is it best to provide a list of discrete behaviors and depend upon the teachers' interpretations? Should the researcher place the interpretive screen on his data and provide the practitioner with more general behavior patterns? Is it best to share theoretical speculations with the teacher and assume that he/she knows the classroom and can best determine what is optimal behavior?

Whether the researcher provides discrete behavioral directions such as "smile," "touch," and "ask questions," suggests a more generalized pattern of behavior such as "warm enthusiasm," teaches a theoretical notion similar to "an attitude of acceptance and interest on the part of the teacher is associated with ...," or offers a combination of these strategies is an important issue in the dissemination of research findings in teacher education programs. It is important that teachers not only be aware of, but also be able to use knowledge gained through careful, thoughtful investigation.

Not all of the participants agreed with the either/or-ness of the high vs. low-inference discussion. One participant maintained that both high and low levels of abstraction were desirable in research. He suggested that the high-inference variables could provide the conceptual structure for the interpretation of the low-inference variables. In this respect then, both levels of abstraction are necessary for both comprehensive investigation and dissemination of findings.

The discussion of the problems in efficient and effective accumulation of evidence also included the issue of the diverse and often undefined methods used in research investigations. The variety of instruments, metrics, outcome measures, and analysis procedures across investigations

coupled with differing relationships between independent and dependent variables and differing units of analysis constitute an accumulator's nightmare. In the Context Variables session of the Invisible College meetings, concern was expressed over the diverse operational definitions and metrics employed in various research efforts. For example, where one researcher may use frequency or percent of behaviors, another may employ percent of time engaged in behaviors while a third may use factors derived from a factor analysis of observed process variables. Clearly needed is a serious effort to make one's definitions and methods understood. Perhaps even more desirable would be a systematic effort to employ common definitions and methods across investigations and investigators.

Strategies Of Accumulating Evidence

One purpose of accumulating evidence is to develop summary statements describing the relationship among variables from independently conducted research endeavors. Examples of such statements include: "Teacher indirectedness is unrelated to pupil achievement" (15 correlational studies are cited in Dunkin and Biddle, 1974, p. 115); "...there is a highly significant relationship between teacher indirectedness and student achievement" (Gage, 1976, p. 14); "..on the average, the therapy group mean was about two-thirds standard deviation above the control group mean on the outcome variable" (Glass, 1976, p. 7).

A number of the systematic strategies are available for accumulating findings to derive such conclusions. Three major strategies of accumulation were mentioned and each was evaluated in terms of the metric of accumulation and the meaningfulness of the resulting summary. Other strategies that were not discussed at the session are presented in Rosenthal (1971), Light and Smith (1971), and Jones and Fiske (1953).

The first strategy evaluated was vote counting, one variation of which is exemplified by the voluminous review of research on teaching by Dunkin and Biddle (1974). This variation of the vote count involves calculating the number of times a relationship is reported in the literature as statistically significant and comparing this total to the number of times the relationship does not reach significance. In evaluating this procedure one participant noted that by reviewing research on the basis of statistical significance the accumulator may risk commiting a Type II error. That is, the accumulator may conclude that no relationship may be present. The following reasoning was offered: First, the true relationship between any single teaching variable (e.g., indirectedness) and student outcomes will probably be of low magnitude. Because of the complexity of the teaching-learning situation, the variable would account for only a small proportion of variance in the dependent measures. Second, research on teaching is usually conducted with a small sample of teachers. Third, to obtain statistical significance with a small sample, the magnitude of the observed correlation must be fairly large. In sum, because of small sample sizes and the low true correlation, the relationship among variables may not reach significance and the resulting box score will not adequately reflect teaching effects which may exist (Gage, 1976). During the dialogue participants observed that in accumulating evidence from professional reports the number of negative findings may actually be underestimated due to the "wastebasket effect" -- throwing out or not publishing negative results.

A second strategy, testing the significance of combined results (Jones and Fiske, 1953), was then offered as an alternative to the vote count. This method tests whether a composite probability derived from separate probability levels of each study is significant. Rather than

categorizing results as above or below a critical level, the numerical probability level of the original test is preserved as the metric of accumulation. These values are converted to an overall $X^2$ chi square which is then assigned a probability from the appropriate statistical table. This composite probability indicates, then, the significance of the relationship between the variables of separate studies.

Concern was expressed, however, that this procedure does not permit the accumulator to derive meaningful conclusions. It was pointed out that by using statistical probability levels as the metric, the accumulator ignores the strength of relationship between variables in the original study. A relationship between two variables may be statistically significant, indicating that the relationship may be replicated reliably; however, the strength or magnitude of the relationship may not be great enough to have a meaningful impact in an educational setting. One discussant observed that a researcher, calculating a correlation coefficient on data from a large sample of teachers, is "doomed to significance" even though the relationship may not be meaningful. This same critique also applies to the derived combined probability. The combined probability does not specify the magnitude of the relationship between the variables across all studies. In addition, the combined probability is influenced by the sample size; when a larger number of studies are combined, the derived probability will be smaller, and hence, significant. For these reasons it was strongly recommended that the metric of accumulation reflect the strength of the initial relation such that a measure reflecting the combined magnitude of the relationship may be derived.

In addition, it was recommended that evidence be combined only within

configurations of contextual and individual difference variables. One discussant noted that the absence of teacher effects in accumulated research may be due to the practice of accumulating across contexts and individual-difference variables. A particular teaching decision, behavior, or strategy may be differentially effective for different lesson formats, ability levels of students, grade levels, etc. If all research on a specific teacher behavior or decision is accumulated in one sum these interactive relationships may be masked in the resulting summary. Also, accumulation within contextual or individual difference variables can facilitate more accurate generalizations to the appropriate populations of subjects and settings. (See proceedings of the session on Context Variables for a more detailed analysis of this topic.)

The third major strategy discussed was the meta-analysis approach used by Gene Glass (1977, in press). Glass's procedure 1) preserves the size of the effect in the original study as the accumulation metric, 2) derives a summary based on a descriptive statistic permitting interpretation of the size of the overall effect, and 3) accounts for the contextual and individual difference variables in the different studies. Glass's metric of accumulation is the difference between the mean of the experimental group and the mean of the control group divided by the within-group standard deviation. This unit is then averaged across all investigations, and the resulting index reflects the average size of the experimental effect in standard deviation units. Glass also subdivides the original studies according to contextual and individual difference variables. For example, in accumulating evidence of the effects of different psychotherapy treatments, all comparisons are made between groups that are equivalent on such factors as duration of therapy, type

of complaint, etc. Thus, this accumulation procedure is less likely to mask interactive effects and would permit more appropriate generalizations of the conclusions.

The discussion of alternative strategies raised several general concerns regarding the feasibility of accumulating evidence from research on teaching at this time. How many studies are required to accumulate evidence within complex configurations of contextual and individual difference variables? Is it possible that there are so many higher order interactions in research on teaching that accumulation is not feasible? Is the research so diverse and uncoordinated that accumulation within configurations is not presently possible despite the availability of a large body of evidence? In order to accumulate evidence within complex configurations, how similar should studies be, and in what respects? Clearly, the lack of systematic and coordinated research is a central problem of accumulation. Discussion of, and reactions to this problem are addressed in the final section of this report.

Unit of Accumulation

Opinions also differed in regard to the optimal unit for accumulation of findings and whether or not accumulation can appropriately cut across different units. Three major units were defined and discussed: variable, context, and perception. This issue of a meaningful unit was discussed in terms of the appropriate unit for any research investigation as well as that unit which is best accumulated. For each group of discussants the favored unit of accumulation was identical to the preferred research focus, and these two issues were frequently confounded.

Some participants advocated accumulation at the variable level. Whether the variable is low or high-inference, these researchers felt that

behavior can be meaningfully compared across investigations and that the appropriate unit for the accumulator to study is the process variable (i.e., those variables which describe what the teacher and students do). For many investigators these process variables are best understood within the framework provided by both demographic variables (SES, age, sex, etc.) and subject matter variables (math, reading, science, etc.). The primary concern, then, is that variables used in investigations be clearly defined and systematically measured.

In contrast to this position it was argued that no meaningful interpretation of process variables can occur either within a study or across studies without attending to the ecological context within which the behavior occurs. Some discussants maintained the now classic stance that the context or behavior setting -- which is observable, has specific props and a signal system, and therefore demands a particular pattern of behavior while precluding others -- is the major determinant of behavior. This position holds that behavior both occurs in and is determined by the context, and without knowledge of the context, one cannot determine the appropriateness of the observed behaviors. This conception of context or behavior setting contrasts with the mere noting of relevant demographic or subject matter variables referred to above. Context, then, is more than a listing of subject and classroom descriptors. An example was given of the significant differences between teacher and student behavior which occurs in the reading group as opposed to the art circle. It seems reasonable, then, that the behavior change observed cannot be understood and correctly interpreted without acknowledging the context within which the behaviors occurred. It is the context (i.e., the reading group or the art circle) which accounts for the significant change in behaviors (Kounin and Gump, 1974).

One participant stated that the distinction between variable as the research or accumulation unit and context as the research or accumulation unit was inappropriate. He maintained that behavioral variables and contextual notions are intertranslatable. Thus, one need not be concerned with careful monitoring of the context because those aspects of the context that are important are translated into the behavior. Any behavior that occurs in the classroom, then, both incorporates and translates the signals and constraints of the environment; to distinguish between and attend to the behavioral variables and the context is therefore redundant and unnecessary. This position proved to be controversial.

Still another position was proposed by those who advocate individual perception as the optimal unit for study and accumulation. It was suggested that the subjective perceptions of the teacher and students are of most value to research. This position surfaced when the participants were generating possible explanations for a mismatch between theoretical expectations and actual research findings. It was posited that a mismatch could occur because of the subjective perceptions of the participants. Thus, students may learn material presented them not because of its organization or style of delivery, but because they suspect it will occur on a test. In other words, it may be less a matter of optimal "chunking" by the teacher and more a matter of speculation about future test material that causes the student to learn. In order to understand a phenomenon in a meaningful way, then, the researcher must investigate how the phenomenon is perceived by the participants in the situation. The researcher cannot project his own perceptions upon the subjects anymore than he can deny that perceptions exist. Here, neither the data collected through

experimental observation nor the context within which the data were collected are as important as the perceptions of the participants which can transform ostensibly different events into functionally equivalent phenomena (Doyle, 1977; Walberg, 1977).

These positions concerning the appropriate unit for investigation and the effective accumulation of investigations were vigorously discussed. Obviously, the unit chosen for accumulation will have a significant effect on the final interpretation of individual research findings because the ultimate conclusions derived from any one of the three different strategies can differ radically from the others. These issues, as well as the possibility of accumulation across units, need to be given considerable thought and energy if meaningful accumulation is to occur.

Issues In Weighing Evidence

The selection of the unit of accumulation is a conscious and deliberate choice made by the accumulator, and the resulting review article or intervention program will undoubtedly reflect the chosen perspective. In contrast, the process of weighing evidence involves subtle decisions which guide the accumulation process and shape the final product but are not always evident to the consumer. Two types of decisions in weighing evidence were discussed: (1) selecting or discounting data and (2) weighing evidence from different research perspectives.

When a researcher or accumulator surveys available evidence he selects certain findings and discounts or ignores others. The pivotal issue of the discussion was whether the accumulator's conceptualization or whether the selection is based on actual data. The discussants repeatedly reported that conceptualizations and assumptions play a paramount role in reviewing

findings. The accumulator is likely to select evidence which supports his conceptual framework or for which he can find a rational explanation. Conversely, incongruent or unexpected evidence may be discounted or ignored.

Participants then discussed two types of bias which influence selectively the accumulation process. One organizational scheme or source of bias is the theory, philosophy, or conception of the teaching-learning process supported by the accumulator. As an example of discounting influenced by a philosophy of teaching, consider a reviewer examining the results of Wright's (1975) study of open classrooms. Wright's research indicated that anxiety of students in open classrooms was higher than that of students in traditional classrooms. This initial and unanticipated result is contrary to the goals and philosophy of most open environments. The reviewer may disregard these results by stating that the instrument measuring anxiety was inappropriate for open classrooms.

A second organizational framework, and probably the more influential bias, is the accumulator's knowledge of trends in previous research. As one participant noted, conceptualizations are not rigid but change primarily through knowledge of recently published research. For instance, when additional evidence indicates that student anxiety is greater in open than traditional classrooms, the accumulator will attend to these once anomalous findings, incorporate them into his conceptual framework, and subsequently accumulate further evidence on this topic.

During the course of discussion it was implied that the accumulator's or researcher's conceptual framework will influence the rate of development of conceptions of teaching held by professionals in the field. Evidence contrary to established conceptions or research findings may be

discounted by the researcher or accumulator and attributed to such factors as methodological flaws in the original study. These studies, if replicated at all, would only be repeated after modifications were made in design, instrumentation, or analysis to examine, again, whether the expected relationship exists. Less likely is the possibility that the researcher will reconceptualize the research question or the relationship among variables in the teaching-learning process and design a new study based on these reconceptualizations. Therefore, numerous studies confirming an anomaly will be required to change the conceptions of the individual accumulator or researcher and, ultimately, the body of knowledge about teaching held by researchers in general.

The preceding discussion emphasized how conceptions of teaching and learning will restrict the range and type of evidence that is accumulated. In addition, the accumulator's assumptions regarding the purposes and validity of different research strategies will also structure the accumulation process.

It was suggested that evidence be accumulated according to a hierarchy of research strategies. The proposed hierarchy is composed of ethnographic and observational studies at the base, followed by correlational studies, quasi-experiments, and the true experiment at the apex. Evidence is accumulated within levels of the hierarchy for purposes specific to each level. Descriptions of classroom life would be developed by accumulating within the horizontal level of observational studies. To provide conclusive evidence of the causal relationship between two variables, the accumulator would compile data at the apex. Hints provided by observational and correlational studies are tested through true experiments.

As an alternative to the idea of research strategy hierarchies one

participant proposed that different research approaches be considered as separate and independent indicators of the same phenomenon. The conception of multiple research modes as equally valid implies that accumulation and validation of evidence is made across all modes rather than successively higher steps in the hierarchy. This proposal is based on the recognition of difficulties inherent in conducting naturalistic experiments in the social sciences.

Participants had the opportunity to explore their conceptions of multiple research approaches during a discussion of the Stanford experiment (Program on Teaching Effectiveness SCRDT, 1976). This study tested the validity of the relationships between variables derived from past correlational evidence through an experiment. In the study, teachers were taught to use instructional strategies for a recitation which demonstrated high or low emphasis on each of three teaching variables. The teachers taught lessons on ecology to sixth grade students according to the prescribed strategies, and an assessment was made of the students' immediate and delayed retention and affect. Structuring of the recitation was included as a teaching variable because past correlational research indicated that structure had a positive effect on student learning. Contrary to the researchers' expectations, however, the results of this experiment indicated that there was no difference between the structured and unstructured recitation on the immediate cognitive multiple-choice test.

If the conception of a research hierarchy was in operation, the results of this experiment would be evaluated independently of the correlational evidence and accepted as the real test and primary indicator of the relationship between structuring and achievement. On the other hand, if the different modes of research were perceived as multiple indicators,

the results of the experiment would be compared with evidence from the other modes, and the results of the correlational studies could provide an indication of the true relationship. After reviewing the unexpected and negative findings of the Stanford experiment discussants reaffirmed the importance of structuring by citing the past correlational research and common sense beliefs, thus demonstrating support for the conception of multiple research indicators.

At the conclusion of the discussion many participants seemed to agree that accumulating evidence across research modes was both desirable and necessary in conducting research on teaching. However, it again appears that prevailing conceptualizations about teaching may influence validation across research modes. Did participants support the concept of structuring because of the strength of past correlational research or because of a conviction that structuring was important? The discussion did not include the possibility of having to weigh evidence from different approaches when findings are inconsistent across modes and conceptualizations of the relationship among teaching variables are not fully established. When confronted with this more ambiguous situation, what should the accumulator or researcher conclude from the research and recommend for future directions?

This issue prompts further questions regarding the role of research on teaching vis-a-vis teacher education. On what type of evidence should teaching prescriptions be based? One participant, recognizing that correlational and experimental evidence did not adequately capture what she observed in the classroom, developed some prescriptions based on hunches from her observations. The researcher specifically identified those prescriptions she derived from hunches, as well as those derived from correlational and experimental studies, thus permitting the consumer to

judge the credibility of her evidence. In designing teacher education programs should prescriptions based on observational and correlational studies be implemented directly? Or should prescriptions be implemented only after experimental or naturalistic intervention studies have been conducted? What if additional experiments are tried, but fail to support the favored prescription?

Ways To Plan For Accumulating Evidence

Throughout the discussion it became apparent that there are numerous obstacles to accumulating evidence in the field of research on teaching. Participants described four possible ways that they, as researchers, could facilitate future accumulation efforts.

One issue critical to future accumulation is the development of theories or models of teaching. Some discussants argued that the field of research on teaching is not barren of theories. Through conducting investigations of their own and through knowledge of others' studies researchers have slowly generated informal conceptualizations which they use as organizational frameworks. One participant, however, held a pessimistic view of this situation, stating that accumulation is senseless unless educators move beyond informal conceptualizations and develop theories or models of teaching. A concerted attempt to accumulate evidence will formalize these conceptions. The resulting theory or model would both direct future accumulation and provide guidance to programs of coordinated research.

A second series of suggestions called for coordinating research by developing commonalities across different studies. A number of discussants proposed that researchers list the variables they feel should have high priority for inclusion in future investigation. (In the past, researchers

primarily interested in the relation between process variables and outcomes have contributed to developing such a list. According to Marliave (1976), there was little agreement, however, on the selected variables, even among those within this research perspective. Participants at the Invisible College meetings represented diverse research paradigms, and lack of consensus on this issue was vividly demonstrated in the ensuing discussion. There was disagreement on whether or not it was desirable to accumulate evidence on the basis of variables. When one participant proposed a list of possible common variables (e.g., student initiation, grouping, businesslike, seatwork, management, etc.), other discussants expressed concern that emphasis on such discrete teacher and student variables would ignore ecological or environmental influences. The diversity of opinion on specification of common variables was also reflected in the earlier discussion on a unit of accumulation.

Agreement on common variables was not reached. It was evident from the discussion that agreement on important variables is more likely to occur among those within a common research paradigm than among those from different perspectives. It is questionable whether it is possible or even desirable for researchers from the different perspectives to select common variables to investigate. Perhaps the greatest agreement that can occur among groups representing the different paradigms is for these groups to converge to investigate a common problem or research site, each from its own perspective. The data gathered through the different perspectives would be complementary, focusing on different aspects of a broad problem, with each aspect uniquely and appropriately examined via the given perspective. In any case, before commonalities of

any type are incorporated into future studies, researchers need to engage in extensive dialogue both between and within the different paradigms.

Closely related to the request for preferred variables was the concern for development of similar operational definitions of selected variables with corresponding common metrics of observation and analysis. As decisions on operational definitions can only be made subsequent to the specification of common variables, this concern was only briefly mentioned.

The third recommendation was to develop more systematic and detailed reporting of research results. Some discussants indicated that not all reports included essential information, such as means and standard deviations. On a broader level it would seem essential to provide detailed and explicit descriptions of the following: population, sample, context or behavior setting, program or treatment, and instrumentation. While implying that such detailed specification is desirable for accumulation purposes, the participants did not discuss whether new standards for publication requiring greater specificity ought to be imposed. Would the guidelines for publications be applicable to a wide range of research approaches? Would such guidelines restrict the type of studies generated and ultimately disseminated?

The final recommendation was the continuation of deliberations among members of the research community as exemplified by the Invisible College meetings. One specific suggestion was for subgroups of researchers to tackle issues which surfaced during the meeting but were left unresolved. These smaller working groups would then periodically report their progress to the total group. Another suggestion was for increasing communication

among researchers through newsletters or bulletins. These forms of communication have been used effectively in other disciplines to keep members of the research community informed, to coordinate diverse research programs, and to stimulate new questions and directions for research.

Considering the enormity of the task of accumulating evidence, deliberations such as these are perhaps one way to begin to plan for future accumulation efforts. If accumulation of evidence is a high priority, it is evident that certain modifications have to be made in the current manner of conducting scholarly research. While participants did suggest a number of ways to plan for accumulation, the question remains whether researchers are willing to pursue these or other suggestions and modify their preferred conceptions, methods, and approaches to facilitate the accumulation process.

Bibliography


Doyle, W. Paradigms for research on teacher effectiveness. In L.S. Shulman (Ed.), Review of research in education (Vol.5). Itasco, Ill.: Peacock, in press.

Dunkin, M.J., & Biddle, B.J. The study of teaching. New York: Holt, Rinehart and Winston, Inc., 1974.

Gage, N.L. Four cheers for research on teaching. Paper presented at American Educational Research Association Annual Meeting, San Francisco, April, 1976.

Glass, G.V. Integrating findings: The meta-analysis of research. In L.S. Shulman (Ed.), Review of research in education (Vol. 5). Itasca, Ill.: F.E. Peacock, in press.

Glass, G.V. Primary, secondary, and meta-Analysis of research. Educational Researcher, 1976, 5, 3-8.

Jones, L.V., & Fiske, D.W. Models for testing the significance of combined results. Psychological Bulletin, 1953, 50, 375-382.

Kounin, J.S., & Gump, P.V. Signal systems of lesson settings and the task-related behavior of preschool children. Journal of Educational Psychology, 1974, 66, 554-562.

Light, R.J., & Smith, P.V. Accumulating evidence: Procedures for resolving contradictions among different research studies. Harvard Educational Review, 1971, 41, 429-471.

Marliave, R. Observable classroom variables (Tech. Rep. I-2). Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development July, 1976.

Program on Teaching Effectiveness, SCRDT. A factorially designed experiment on teacher structuring, soliciting, and reacting (Research and Development Memorandum #147). Stanford: Stanford Center for Research and Development in Teaching, 1976.

Rosenshine, B., & Martin, M. Teacher education and teaching behavior: Comments on the state-of-the-research. Educational Researcher, 1974, 3, 11-12.

Rosenthal, R. Teacher expectations and pupil learning. In R.D. Strom (Ed.), Teachers and the learning process. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971, 33-60.

Walberg, H. Psychology of learning environments: Behavioral, structural, and perceptual. In L.S. Shulman (Ed.), Review of research in education (Vol. 4). Itasca, Ill.: F.E. Peacock, 1977.

Wright, R.J. The affective and cognitive consequences of an open education elementary school. American Educational Research Journal, 1975, 12, 449-464.

# OUTCOME MEASURES

by Steve Olejnik and Susan Wildfong

## Introduction and Rationale

Every definition of effective teaching ultimately depends upon stipulating the effects that are to be attributed to the act of teaching. Yet experienced and thoughtful researchers on teaching have invested hours, days, and years in designing powerful treatments with insufficient effort in the identification or construction of student outcomes that are sensitive to good instruction.

The intent of the final session of the Invisible College meetings was to offer a forum from which common problems in measuring teacher effects could be attacked. Efforts to measure these effects have frequently been limited to standardized tests of achievement. The validity of these tests, as well as their sensitivity to important teacher-affected outcomes, has been increasingly questioned. Many outcomes are outside the domain of standardized tests, particularly social-emotional outcomes and higher cognitive skills. In addition, better and more fully understood methods for measuring change in studies of natural variation are needed for all types of outcome dimensions. The goal of the Outcomes session was to bring together researchers in a common effort to resolve these problems. It was hoped that through the sharing of methodology and instrumentation the field of research on teaching would be able to move beyond the present state to more accurate studies of the range of teacher-affected outcomes.

## Overview Of Current Research On Teacher Effects

Three recent studies were presented by their principle investigators during the Outcomes session. These studies provided the framework for a

discussion of several critical problems associated with outcome measures in research on teaching. In outlining each investigation the researchers addressed four general concerns: 1) the context in which the study was conducted; 2) the variables which were of interest and why; 3) the assessment procedures used; and 4) the particular problems which were encountered. In addition to providing a guideline for the discussion, the presentation of the research efforts gave an overview of the current status of research on teaching.

While only three studies were considered at the meeting, the range of the responses to the above concerns varied widely. Settings ranged from preschool day care centers and home to junior and senior high school classrooms. Children as young as 12 months and as old as 18 years were included as data sources for estimating the effects of teacher behavior.

The variables considered were as varied as the settings. While cognitive skills seem to have been the greatest concern across all studies, effects in the affective domain including attitudes and social-emotional development were also seen as important areas for study. The rationale for choosing variables was provided by several factors. Some variables, such as social-emotional skills, were selected because of the ages of the children and the setting of the investigation (e.g., day care centers). Selection of variables from the affective domain was based partially on a feeling of responsibility to address this domain since it includes some of the major goals of education, and partially out of a concern that cognitive gains may be purchased at the expense of affective losses. The values of the researcher were also a major determinant of which variables were selected for. Finally, the ability to measure variables played a large role in selection.

Investigators discussed both the context of their studies and the variables which were of particular interest, then briefly outlined his/her procedure for assessment. As may have been expected, the measurement procedures used were dependent on the variables studied and the settings. The study of very young children used observational techniques in semi-structured situations, while older children were assessed with paper and pencil tests emphasizing cognitive skills. The use of published instruments in assessment played only a minor role in all three studies reviewed. Each investigating team developed its own instruments for measuring either cognitive or affective skills of specific interest to the researchers.

The lack of appropriate assessment techniques was a problem noted by all of the investigators. In particular, the lack of instrumentation for assessing young children was viewed as an important limitation on research efforts. Other problems involved choosing the appropriate unit of measurement from an interpretable point of view, determining the appropriate frequency of observation, and measuring change in a naturalistic setting. Concerns about these problems were held in varying degrees by all the investigators.

It became clear during the discussion that concern for teacher effects encompasses diverse contexts, variables, and measurement procedures. The only common bond between studies appeared to be the problems encountered. These problems and issues provided the context in which further discussion developed. The following sections of this report summarize the major issues raised and suggestions on how they may be resolved.

## Problems and Issues

Choosing outcomes. The first issue addressed was the selection of

outcome variables in research on teaching. It was apparent from the review of the three studies that a number of considerations entered into this decision-making process. The goals and values of the researcher usually surfaced as major contributing factors, while teachers' goals usually played a minor role. This point raises an interesting question concerning the relationship between teacher objectives for a class and the variables that are chosen for investigation. If the classroom goals are adequately reflected in the variables studied, then the effects of teacher behavior can be estimated. But if there is a conflict between the teacher's and the researcher's goals, then what is estimated may not be an adequate assessment of the true effect of the teacher's behavior.

In addition to the potential conflict between teacher and researcher goals, there is also the possibility that conflicts exist in objectives among teachers. For example, a unit studied over a two-week period may consist of four major objectives. Some teachers may concentrate on one or two objectives while others spread their instructional time evenly over the four objectives. A researcher studying teacher behavior during this unit may assess the students on their achievement across all four objectives or may study each objective separately. No matter which approach is taken, the effects of the behavior of some teachers will be misrepresented. Thus, results that have been interpreted as data on teacher ineffectiveness may actually just indicate a discrepancy between the teacher's goals and those reflected in the measures.

Although some critics might question the appropriateness of a comparison of outcomes when teachers have different classroom goals, several solutions to the dilemma were suggested. One possibility might be to include multiple sets of outcome measures, each reflecting a different set of goals. For example, an instrument might be developed that adequately

assesses each individual objective and sums them together to provide an overall estimate of learning the entire unit. This approach, however, may be inadequate in some situations since there are practical limitations as to the length of an instrument and the number of instruments which can be administered to any one group. (This issue is discussed in more detail later in this report).

Another solution suggested was to include in the study only those teachers who share a common set of goals and objectives. The task of identifying this type of homogeneous group might be accomplished by a number of procedures, including interviews to obtain both student and teacher perceptions of goals, the use of teacher logs, and a careful review of teacher-made tests. Use of teacher tests, however, was criticized as invalid since what is taught and what is tested are frequently different. Nevertheless, attempts to insure a commonality of objectives were seen as important considerations for future research on effective teacher behavior.

Closely related to the issue of teacher-researcher goal conflict is the question of motivation. A considerable amount of the literature indicates a strong, positive relationship between level of motivation and test performance. If the outcomes chosen for study do not reflect the objectives of the teacher, low motivation and therefore low test performances may result. This low performance may then lead to underestimating the teacher's effects. Motivation may also be reduced due to social pressures and attitudes. For example, the Association of Black Psychologists and the NAACP have recently called for a moritorium on all standardized testing. To avoid problems of motivation and obtain an accurate estimate of teacher effects it was suggested that researchers provide the

participants with a rationale for using published tests or develop new assessment procedures which may be intrinsically motivating.

Selecting outcomes to investigate teacher effects may also depend to a considerable extent on the length of time over which the study is conducted. A set of outcomes that are relatively easy to modify might be selected for an investigation that spans a two-week period. On the other hand, a completely different set of variables might be chosen if the study were to be conducted over a two-year period of time. This relationship between period of investigation and outcomes was of great concern to several participants of the session. The problem arises partly from current conditions in research on teaching, such as researchers designing experimental studies rather than observational investigations. Since it is extremely difficult to conduct an experimental study over an extended period of time, the variables which are being selected for study are those which are likely to show an effect quickly. Instruments for outcomes such as attitudes and higher order cognitive skills (both of which take greater time to develop) have not been given the kind of study required for adequate development.

Assessing outcomes. A second major issue addressed in the outcomes session was the assessment of teacher effects. As indicated earlier in this report, the assessment procedures used in the three research efforts varied greatly, depending on the variables studied. The observational approach, for example, was adopted in one study because the variables of interest did not lend themselves to other methods of assessment. This study's use of observational techniques created problems, however, when it became apparent that the naturally occurring frequency of the behavior of interest was too low to be useful and a modification of the environment

was required. As a result, the investigators constructed a semi-structured situation which increased the probability that the behavior would occur. In relating their experience with assessment these researchers provided renewed awareness of an important limitation associated with observational techniques. That is, in the study of naturally varying behavior, observational procedures may be appropriate only in situations where the frequency of the behavior of interest is reasonably high.

A second assessment technique mentioned in the discussion of the research studies was the use of paper and pencil tests and questionnaires. Such a technique differes from observational procedures by providing a very specific framework in which behavior is recorded. At the same time, however, paper and pencil instruments share with the observational approach a flexibility to assess variables in both the cognitive and affective domains. The primary source of materials for this approach is standardized published instruments. There are several reasons for this: they have reported psychometric properties; they facilitate study replication and the accumulation of findings; and they reduce developmental time and money.

While these are strong arguments for using published instruments, there are also several limitations associated with this approach. One argument is that standardized instruments do not exist for all of the outcomes that interest researchers on teacher effects. An example was provided in the study of day-care centers and homes. The quality of care children receive is expected to be related to how children grow socially and emotionally, but there are few or no satisfactory standardized tests available to examine a preschool child's self-esteem or social ability.

A second problem with standardized tests involves content. Frequently,

researchers on teaching are concerned with very specific subject areas or materials which could be learned over relatively short periods of time. Standardized instruments do not generally lend themselves to such specific interests. In short, standardized instruments were primarily developed to examine longer-term growth. To obtain a sensitive measure of short-term effects requires the researcher to develop tests that cover the specific content area or materials of concern.

Two of the studies reviewed took this approach. In one study, considerable time and effort were spent designing and developing a multiple choice test that reflected the contents of the unit taught and measured both higher and lower psychological processes. In the other study, the investigators developed criterion-referenced tests in both math and English after an extensive review of course materials and observational and interview data. In addition, both studies developed attitude questionnaires to assess affective outcomes.

While standardized or researcher-developed instrumentation provides a convenient approach to assessing teacher effects, these procedures are severely hindered by the number of tests which can be administered. There are practical limits to the amount of time and effort teachers and students can contribute to test-taking activities. At one point in the discussion it was estimated that, at a maximum, a total of six hours might be afforded to outside testing. Earlier in this report, when the goals and values of teachers and researchers were discussed, it was suggested that the goals of the teacher ought to be taken into consideration when assessing teacher behavior. This may be a laudable objective, but if the goals of a teacher are widely discrepant from the goals of the researcher the amount of testing which would be required to take both sets of goals into consideration might be prohibitive.

Another solution to the problems of time constraint, motivation, and the assessment of higher cognitive skills was offered in the use of teacher-designed projects that call for the application of particular skills of interest. One researcher provided the following example of this type of assessment procedure. A teacher who had been teaching a unit of decimals and percentages asked his students to compute and compare the cost of their lunches for a field trip they were planning if they: 1) brought their own lunch, 2) bought a lunch on the train, and 3) if they bought their lunch at the museum they planned to visit. The reaction from the children was enthusiastic. They were motivated to complete the task since it was relevant to their future plans and they enjoyed doing it. For the teacher the task provided the opportunity to determine how effective his efforts in teaching the unit had been. Reaction to this approach was positive, and participants were hopeful that assessment procedures of this type would be further developed.

A final consideration in assessing teacher effects involved the use of student and teacher perceptions. For the most part, teacher effectiveness studies have relied on observational and paper and pencil tests to understand the complex set of interactions that occur in the classroom. It was pointed out, however, that these interactions might be better understood if examined from the student and teacher points of view. By obtaining the perceptions of the principal actors in the classroom setting greater insight might be gained from the behavior under study. It was therefore suggested that in future research on teacher effectiveness some effort be given to obtaining data on the perceptions of the principal actors in the study.

Measuring change. The final major concern raised in discussion was the problem of measuring change. This issue has plagued educational

researchers and social scientists for many years and continues to be a critical problem for researchers on teacher effects. The problem stems from two independent factors. First, the effects of teacher behavior are based on changes in student performance that occur during the period of study. Secondly, research on teacher effects has generally tended to be a study of natural variation as opposed to intervention. To estimate the effects of teacher behavior, then, the investigators have been forced to adjust for initial differences in the performance levels of the students and have assumed that no other differences develop during the study that are not attributable to teachers.

A number of adjusting procedures have been suggested and used, ranging from simple raw gain scores to relatively complex residualized gains. The perils of such adjustments, however, are pointed out in a number of discussions in the literature (Campbell and Erlebacher, 1970).

As an example of one such problem consider a study of the relationship between achievement and the type of questions a teacher asks. Rather than conducting an experiment where the types of questions are manipulated, a frequent approach is to conduct a study of naturally varying behavior. A group of teachers who have a history (previously observed over several years) of asking higher order questions are identified. Similarly, a group of teachers who tend to ask lower order questions are also identified. A researcher may then look at the achievement gains of the students of each group to determine whether the type of questions asked has any effect on achievement. It may be true that the teachers differ significantly in other respects, but for this example it is assumed that there were no other obvious differences.

After obtaining a pretest on all of the students the researcher finds that teachers who tend to ask higher order questions are instructing students

who, on the average, have a higher initial achievement score than those students taught by the teachers who ask lower order questions. After obtaining the student post-test scores the investigator adjusts for the initial differences with any of a number of residualizing procedures. The results then indicate that the group of students taught by the teachers asking higher order questions achieved more than the other group.

If in fact there were no actual teacher effects, the investigator may conclude either that students who start off high tend to gain more, or that the teachers who ask higher order questions are better teachers. While the former is the appropriate conclusion--the rich get richer phenomenon -- the latter, inappropriate conclusion is the one likely to be accepted since it fits into our conception of ideal teaching behavior. The problem here is actually one of differential growth rates. We do not really know how children grow; consequently, the selection of the adjustment procedure is based only on a guess as to the appropriate growth model. Interventions do not suffer from this problem since randomization in the long run equates growth rates for treatment groups. At present there is no single solution to the problem, but a recent paper by Bryk and Weisberg (1977) addresses the issue.

For now, researchers studying natural variation continue to struggle with the problem of measuring change. A solution suggested in one study was to analyze the data using several different methods of residualizing gains. The results of that study indicated a similar conclusion across all methods considered. It was then suggested that initial performance be correlated with each procedure of calculating residual gains. If the correlation is low and nonsignificant, it may be concluded that the appropriate adjustment was made. On the other hand, if the correlations were

significant it seems unlikely that the measure of gain reflects changes brought about by the teacher's behavior, but rather that the adjustment procedure is inadequate in removing all initial differences between students. This inadequacy is caused in part by the procedure researchers frequently use to calculate residualized gain. Generally, the outcome of interest is residualized using student characteristics within a classroom and then averaged to compute the adjusted gain of a class. These adjusted gains are then correlated with some process variable of interest to obtain an estimate of the relationship between the process variable and the adjusted outcome.

A more appropriate procedure, however, is to first compute the mean of the outcome variable and the residualizing variable and then compute the residualized gains across classrooms using the means. This procedure is the better approach since it uses the classroom rather than the student as the unit of analysis. Typically, the slope used in defining residualized gains on individual students is less than the slope used in defining such gains on classroom means; thus, aggregating residualized gains on students will leave a correlation with initial performance.

A related point is that forming residualized gains using an estimated slope treats that estimated slope as a population parameter, which it is not. One can sidestep this problem by entering both process and initial performance variables simultaneously in a regression equation to predict post-test performance using classrooms as the unit of analysis. The hypothesis of interest is then tested by comparing the beta weight for process to zero.

Finally, in the debate on using residualized gains to estimate teacher effects it was pointed out that these procedures may diminish the relationship between the process variable and outcomes. If the characteristics of

the students in a class influence the level of the process variable being studied and these characteristics are the variables on which residualization is based, then in addition to removing initial differences between classes the researcher will also remove the effect of the process variable. Take, for example, the study of the relationship between the type of question a teacher asks and student achievement. If the students in a class influence the type of questions a teacher asks (a class of bright students may encourage a teacher to ask higher order questions), then adjusting for initial differences in class ability will also remove any differences in the frequency of higher order questioning, thus spuriously reducing the relationship between the type of question asked and achievement. To the extent, then, that teacher behavior is confounded with pretreatment variables the relationship between behavior and the measure of change is diminished. Although there is no solution to this problem investigation can estimate the extent of the confounding by the magnitude of the correlation between the initial characteristics and the process variable. If the correlation is low, a residualizing procedure may be appropriate. If it is high, the investigator may conclude that residualization is inappropriate.

Summary

Researchers on teaching are attempting to assess student outcomes in many different dimensions with a variety of instruments while continuing to struggle with the problems of distinguishing and enumerating the true effects of teaching. Four possible suggestions on the direction of future research seemed to emerge from the discussion on outcome measures. First, teacher goals have to be considered in further studies of teacher effects. Second, both observational and experimental procedures are needed to

to measure immediate and long-term effects of teacher acts. Third, multiple assessment procedures are needed to measure all the outcomes that are valued in education. Neither standardized tests nor teacher-made tests are adequate in achieving this goal, but there seems to be promise in the idea of using application tests of particular skills. Finally, greater effort is needed to determine how children actually grow and to develop appropriate analytical strategies that reflect these growth models.

Bibliography

Byk, A.S., & Weisberg, H.I. Use of the nonequivalent control group design when subjects are growing. Psychological Bulletin, in press.

Campbell, D.T., & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Compensatory education a normal debate (Vol. 3), Disadvantaged child. New York: Brumer-Mazel, 1970.

Coleman, J.S., Campbell, E.G., Hobson, C.J., McPortland, J., Mood, A.M., Weinfeld, E.D., & York, R.L. Equality of educational opportunity. Washington, D.C.: Government Printing Office, 1966.

Evertson, C.M., & Brophy, J.E. The Texas Junior High School Study: Overview of methodology and rationale. Paper Presented at the Annual Meeting of the American Educational Research Association, New York, 1977.

Jencks, C. Inequality, a reassessment of the effect of family and schooling in America. New York: Basic, 1972.

Mosteller, F., & Moynihan, D. (Eds.) An equality of educational opportunity. New York: Vintage, 1977.

Program on Teaching Effectiveness. A factorially designed experiment on teacher structuring, soliciting, and reacting (R & D Memorandum No. 147). Stanford: Stanford Center for Research and Development in Teaching, 1976.

Stallings, J., & Kaskowitz, D. Follow through classroom observation evaluation, 1972-1973. SRI Project URU-7370, August, 1974, Stanford Research Institute.

## PAPERS AND REPORTS DISTRIBUTED AT MEETINGS

The following manuscripts and materials distributed among participants of the Invisible College Meeting, November 17-19, 1976, at the Institute for Research on Teaching. Reprints may be obtained by writing directly to the author(s).

### Manuscripts


Brophy, J.E., & Evertson, C.M. Context effects on classroom process variables. (Rep. 76-10). Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1976.

Doyle, W. Learning the classroom environment: An ecological analysis of induction into teaching (DEF). Denton, Texas: North Texas State University, 1976. (mimeographed summary)

Doyle, W. The uses of nonverbal behaviors: Toward an ecological model of classrooms. Merrill-Palmer Quarterly, in press.

Doyle, W., & Ponder, G.A. The practicality ethic in teacher decision making. Denton, Texas: North Texas State University, 1976. (mimeographed)

Elstein, A.S. Clinical judgment: Psychological research and medical practice. Science, 1976, 194, 696-700.

Erlich, O., & Shavelson, R. The search for correlations between measures of teacher behavior and student achievement: Measurement problem, conceptualization problem, or both? Los Angeles, California: University of California. (mimeographed)

Fenstermacher, G.D. The question of quality in teacher education. Address to general session of the fall conference of the California Council on the Education of Teachers, November 5, 1976, at Yosemite National Park, California. (mimeographed)

Gage, N.L. A factorially designed experiment on teacher structuring, soliciting, and reacting. Journal of Teacher Education, 1976, 27, 35-38.

Good, T.L., Grouws, D.A., & Beckerman, T.M. Curriculum pacing: Some empirical data in mathematics. Columbia, Missouri: University of Missouri, 1976. (mimeographed)

Karlson, A.L., & Stodolsky, S.S. Predicting school outcomes from observations of child behavior in classrooms. Paper presented at annual meeting of American Educational Research Association, February 1973, at New Orleans. (mimeographed)

Kounin, J.S., & Doyle, P.H. Degree of continuity of a lesson's signal system and the task involvement of children. Journal of Educational Psychology, 1975, 67, 159-164.

Kounin, J.S., & Gump, P.V. Signal systems of lesson settings and the task-related behavior of preschool children. Journal of Educational Psychology, 1974, 66, 554-562.

Mitman, A.L. The combined use of quantitative methods and qualitative cross-validation in educational research. Stanford, California: Stanford Center for Research and Development in Teaching, Stanford University, September 1976. (mimeographed)

Peterson, P.L. Interactive effects of student anxiety, achievement orientation, and teacher behavior on student achievement and attitude. Stanford, California: Stanford University School of Education, May 1976.

Rosenshine, B. Current topics in research on teaching. Urbana, Illinois: University of Illinois, October 1976. (mimeographed)

Shavelson, R.J., Cadwell, J., & Izu, T. Teachers' sensitivity to the reliability of information in making pedagogical decisions. Los Angeles, California: University of California, 1976. (mimeographed)

Shavelson, R., & Dempsey, N. Generalizability of measures of teacher behavior.

Shulman, L.S., & Elstein, A.S. Studies of problem solving, judgment, and decision making: Implications for educational research. In F. Kerlinger (Ed.), Review of research in education (Vol. 3). Itasca, Illinois: F.E. Peacock Publishers, 1975.

Solomon, D., & Kendall, A.J. Individual characteristics and children's performance in varied educational settings (Summary of final report, Spencer Foundation Project). Rockford, Maryland: Psychological Services Section, Montgomery County Public Schools, 1976. (mimeographed)

Stoldolsky, S.S. Observational studies of variation in child behavior in classrooms. Chicago, Illinois: University of Chicago, 1976. (mimeographed abstract)

Tikunoff, W.J., & Ward, B.A. Effective teacher education program, work scope for 1976: Some selected findings from three studies. San Francisco: Far West Laboratory for Educational Research and Development, 1976. (mimeographed)

Viehoever, K. The National Conference on Studies in Teaching: A summary and analysis of process, structure, and outcomes. Washington, D.C.: National Institute of Education, 1975. (mimeographed)

Viehoever, K. Some issues in social science methodology, with suggested resolutions: From detente toward integration. Washington D.C.: National Institute of Education, August 1976.

Vinsonhaler, J.F., Chan, Sui-Wah, Wagner, C., & Elstein, A.S. Computer simulation of clinical decision making: A new approach. East Lansing, Michigan: Michigan State University, 1976. (mimeographed)

Ward, B.A., & Tikunoff, W.J. An interactive model of research and development in teaching (Rep. A76-1). Teacher Education Division Publication Series. San Francisco, California: Far West Laboratory for Educational Research and Development, 1976.

Miscellaneous

CLIPIR Project Summary, First Operational Year (November 1976) - December 1977). Institute for Research on Teaching, Michigan State University, East Lansing, Michigan, September 15, 1976.

CLIPIR and Closely Related Projects. (Chart)

Teacher Training Packet 1: An Introduction and Classroom Vignette. Program on Teaching Effectiveness, Stanford Center for Research and Development in Teaching, Stanford University, Stanford, California, Fall 1976.

Study of Third-Grade Instruction: Questionnaire. Program on Teaching Effectiveness, Stanford Center for Research and Development in Teaching, Stanford University, Stanford, California, 1976.

Appendix A

# Invisible College Meetings: Evaluation and Analysis*

Kent Viehoever, NIE

## Background and Participants

In August of 1975, persons associated with the California Beginning Teacher Evaluation Study (BTES) organized a meeting of researchers in the field of teacher education at San Diego. The following November, this same group formed the researcher core at a joint meeting of researchers and practitioners in Austin, Texas, dealing with Teacher Effects. The group met again at the San Francisco AERA meetings in April, 1976. The meetings just held at MSU are generally considered to be the fourth such meeting of the group.

The participants originate primarily from the Far West Lab and the Stanford and Texas R&D Centers. The 1974 National Conference was dominated by Stanford, and the Austin conference by the Texas crowd. The prime movers behind the three smaller meetings through July of this year were from the Far West Lab. Now MSU has moved toward a leading role, both by hiring staff from the other institutions and by formalizing the group as an "invisible college" conceptualized as an integral arm of the IRT.

## Discussion of Selected Problem Areas

It was my initial opinion that the four topics were well selected and well described. With the advantage of hindsight, however, it is clear

*Report sent to Virginia Koehler, Tom Sticht, and Frank Sobol of NIE on November 26, 1976.

that (a) despite the provocative questions outlined ahead of time, there was insufficient structuring in the sessions to force a converging focus on the central problem issues, and that (b) there was no provision for relating the problem areas to the papers contributed by those attending the meetings--nor were the authors, in the absence of any such encouragement, inclined to be so presumptuous as to inject into a "group problem-solving" setting their own already worked-out solutions. Thus the sessions tended to be poorly focused and did not capitalize on the talent available. Perhaps this reservoir could have been tapped (without sacrificing the intended informality) if a quick sort of the papers had been made, along with partial readjustment of discussion topics, to allow an identification of the most relevant papers for each topic.

Time. This session was better focused than others, since it started out from Berliner's presentation and drew on related findings from others (e.g., Gage et al.).

Contextual Variables. Extremely poor focus. I presume this testifies to the admirable self-control of participants, inasmuch as at least 10 of the submitted papers dealt with contextual questions! However, authors of those papers may have felt that their definition of context had little contact with the rather circumscribed discussion in this session.

Accumulating Findings. Although this session started out with the apparent intent (a la Rosenshine) of focusing on a set of variables about which there were some converging findings, it became clear that the most vocal participants were giving greater weight to findings which made sense within some coherent rational framework. The stating of this concern was perhaps the single most important outcome from the discussion sessions.

Outcome Measures. This session started out with the questioning of several researchers concerning their choices regarding outcome measures.

I could see no clear "jelling" of this line of probing, but we left before the session was concluded or before Gage started his scheduled Summary.

Distribution of Participants' Recent Papers

In contrast to many conferences which are overstructured in terms of participant papers, these meetings made essentially no use of the approximately 25 papers submitted for distribution at the meetings. Relevant papers could at least have been afforded overview attention in specific sessions. For example, the Outcome Measures session could have benefitted considerably from attention to two of Shavelson's papers, while the Karlson-Stodolsky paper provides a stimulating inversion of the usual questions re outcome measures.

Individual and Small-Group Discussions

Break periods, meals, and cocktail evenings allowed for supplementary discussions. Attendees included a guest from the Netherlands (Peter Karstanje) who is in this country to evaluate the operation of middle schools here. The Netherlands is about to embark on an ambitious program changing the educational system from nursery-school age through higher education and continuing education. The most radical component is the introduction of middle schools throughout for all 12-16 year olds, but the plan also anticipates incorporating the 4-5 year olds' nursery schools into the primary school, extending the high school through 18-19 year olds, and forming linkages to continuing higher education in the "open school" (note to be confused with the term "open school" in this country).

Appendix B

Invited Participants

Invisible College Meetings

Institute for Research on Teaching

November 17-19, 1976

Richard Anderson
Ctr for the Study of Reading
School of Ed.
Univ. of Illinois
Urbana, IL 61801

Thomas Anderson
Ctr for the Study of Reading
School of Ed.
Univ. of Illinois
Urbana, IL 61801

David Berliner
Far West Laboratory
1855 Folsom St.
San Francisco, CA 94103

Gaston Blom
Department of Psychiatry
B101 W. Fee
Michigan State University
East Lansing, MI 48824

Cary Borich
Univ. of Texas
Research and Development Ctr.
for Teaching
Austin, TX

Robert Bridgham
Med. Ed. Res. & Dev.
202E Fee Hall
Michigan State University
East Lansing, MI 48824

Wilbur Brookover
Urban Metro Stds.
442 W. Owen Hall
Michigan State University
East Lansing, MI 48824

Jere Brophy
Inst. for Research on Teaching
213 Erickson Hall
Michigan State University
East Lansing, MI 48824

Byron Brown
Economics
201 Marshall Hall
Michigan State University
East Lansing, MI 48824

Joe Byers
Inst. for Research on Teaching
209 Erickson Hall
Michigan State University
East Lansing, MI 48824

Chris Clark
Inst. for Research on Teaching
208 Erickson Hall
Michigan State University
East Lansing, MI 48824

Diane Colbert
Learning Research Development Ctr.
University of Pittsburgh
Pittsburgh, PA. 15213

John Crawford
Ctr. for Research and Development
in Teaching
Stanford University
Stanford, CA 94305

Phillip Cusick
Admin. and Higher Ed.
409 Erickson Hall
Michigan State University
East Lansing, MI 48824

Walter Doyle
School of Education
North Texas State University
Denton, TX 76203

Gerald Duffy
Elem. Spec. Ed.
201E Erickson Hall
Michigan State University
East Lansing, MI 48824

Arthur Elstein
Med. Ed. Res. & Dev.
217E Fee Hall
Michigan State University
East Lansing, MI 48824

Frederick Erickson
Graduate School of Education
Harvard University
Longfellow Hall, Appian Way
Cambridge, MA 02138

Carolyn Evertson
37 Sundown Parkway
Austin, TX 78746

Julia Falk
Dept. of Linguistics
A608 Wells Hall
Michigan State University
East Lansing, MI 48824

Robert Floden
Inst. for Research on Teaching
205 Erickson Hall
Michigan State University
East Lansing, MI 48824

N. L. Gage
SCRDT
Stanford University
Stanford, CA 94536

Thomas L. Good
University of Missouri
Columbia, MO 65202

Philip Jackson
Dept. of Ed.
5835 Kimbark Avenue
Univ. of Chicago
Chicago, IL 60637

Norman Kagan
Couns. Pers. Services
434 Erickson Hall
Michigan State University
East Lansing, MI 48824

Virginia Koehler
National Inst. of Education
1200 19th St., N.W.
Washington, D.C. 20208

Jacob Kounin
Educational Research Dept.
College of Education
Wayne State University
Detroit, MI 48202

Judith Lanier
Inst. for Research on Teaching
318 Erickson Hall
Michigan State University
East Lansing, MI 48824

Jerry Lanier
Elem. & Spec. Ed.
201C Erickson Hall
Michigan State University
East Lansing, MI 48824

Barry Lezotte
Inst. for Research on Teaching
203 Erickson Hall
Michigan State University
East Lansing, MI 48824

John Lopis
Teacher Education
201B Erickson Hall
Michigan State University
East Lansing, MI 48824

Hermine Marshall
Lecturer in Education
17 Norwood Avenue
Kensington, CA 94707

Fred McDonald
Educational Testing Service
Rosedal Road
Princeton, New Jersey 08540

Connie McIntyre
Elem. and Spec. Ed.
201 Erickson Hall
Michigan State University
East Lansing, MI 48824

Carol Moore
College of Ed.
Ohio State University
19 Woodruff
Columbus, OH 43210

Vreta Morine-Dershimer
Far West Laboratory
1855 Folsom St.
San Francisco, CA 94103

Penelope Peterson
Dept. of Educational Psychology
Univ. of Wisconsin-Madison
1025 W. Johnson Rm 863
Madison, WI 53706

Andrew C. Porter
Inst. for Research on Teaching
207 Erickson Hall
Michigan State University
East Lansing, MI 48824

Majorie Powell
State of California
Comm. for Teacher Prep.
& Licensing
1020 "O" St.
Sacramento, CA 95814

Billie Rader
College of Education
Michigan State University
East Lansing, MI 48824

Laura Roehler
Elem. and Spec. Ed.
215 Erickson Hall
Michigan State University
East Lansing, MI 48824

Barak Rosenshine
College of Education
University of Illinois
Urbana, IL 61801

Joseph Schwab
Ctr. for the Study of Democratic
Institutions
Box 4068
Santa Barbara, CA 93103

William Schmidt
Couns. Pers. Serv.
458 Erickson Hall
Michigan State University
East Lansing, MI 48824

Richard Shavelson
UCLA
Dept. of Education
Los Angeles, CA 90024

Lee Shulman
Inst. for Research on Teaching
461 Erickson Hall
Michigan State University
East Lansing, MI 48824

Curt Smitch
Teacher Ed.
211 Erickson Hall
Michigan State University
East Lansing, MI 48824

Louis Smith
Grad. Inst. of Education
Washington University
St. Louis, MO 63130

Robert S. Soar
513 Weil Hall
Inst. for Dev. of Human Res.
University of Florida
Gainesville, FL 32611

Francis Sobol
Nat. Inst. of Education
1200 19th St., N.W.
Washington, D.C. 20208

Daniel Solomon
Montgomery County Public
Schools
850 Hungerford Dr.
Rockville, MD 20850

Jane Stallings
Stanford Research Inst.
Menlo Park, CA 94025

Thomas Sticht
Assoc. Dir.
Basic Skills Div.
Nat. Inst. of Education
1200 19th St., N.W.
Washington, D.C. 20208

Susan Stodolsky
Dept. of Ed.
Univ. of Chicago
5835 S. Kimbark
Chicago, IL 60637

William Tikunoff
Far West Laboratory
1855 Folsom St.
San Francisco, CA 94103

Kent Viehoever
Nat. Inst. of Education
1200 19th St., N.W.
Washington, D. C. 20208

John Vinsonhaler
Couns. Pers. Serv.
450 Erickson Hall
Michigan State University
East Lansing, MI 48824

Margaret Wang
Learning Research Dev. Ctr.
University of Pittsburgh
Pittsburgh, PA. 15213

Herb Walberg
University of Illinois
at Circle
Chicago, Illinois 60680

Donna Wanous
Teacher Ed.
410 Erickson Hall
Michigan State University
East Lansing, MI 48824

Beatrice Ward
Far West Laboratory
1855 Folsom St.
San Francisco, CA 94103

Rhona Weinstein
Dept. of Psychology
Univ. of California
Berkeley, CA 94720

Appendix C

# AGENDA
## INVISIBLE COLLEGE MEETINGS

WEDNESDAY, November 17

9 - 12 noon, 133 Erickson

Chairpersons: J. Lanier and V. Koehler

Presentations:
1. Welcome and Preview of Agenda
2. Presentation - "Current NIE Plans and Prospects"
3. Report - "Center for the Study of Reading"
4. Review of Work on Medical Decision Making
5. Overview of Current IRT Activities. Report on Clinical Information Processing in Reading Project as an example.

12 noon, Shaw Hall

Lunch. Guests may purchase meal tickets at the West Shaw Reception Desk. Guests will go through the West cafeteria line. Tables have been reserved in the 1949 Room.

1:30 - 4:30 p.m., 133 Erickson

Chairpersons: D. Berliner and C. Clark

Recorders: R. Yinger and T. Harris

Discussion Topic: TIME

4:45 p.m., Lobby, Erickson

Transportation to the University Inn will be provided by IRT staff. Meet in front of Erickson at 4:45 p.m.

7:00 p.m., University Club of Michigan State University

Refreshments 7:00 p.m.; Dinner 7:30 p.m.

Transportation will be provided by the University Inn to the University Club at 6:45, 7:00 and 7:15. Arrangements may be made at the motel reservation desk.

Return transportation to the University Inn will be provided by IRT staff.

THURSDAY, November 18

8:30 - 11:30 a.m., 133 Erickson

Chairpersons: J. Brophy and T. Good

Recorders: J. Passalacqua and J. Elmore

Discussion Topic: CONTEXT VARIABLES

THURSDAY, November 18

11:30 - 1:00 p.m.

Lunch. On your own. (Shaw Hall, International Center, Kellogg Center, MSU Union, etc.)

1:00 - 3:30 p.m., ERICKSON KIVA

Chairpersons: B. Rosenshine and L. Shulman

Recorders: M. Rohrkemper and B. Perles

Discussion Topic: ACCUMULATING FINDINGS

4:00 - 5:15 p.m., ERICKSON KIVA

IRT COLLOQUIUM, D. Berliner

"Beginning Teacher Evaluation Study"

5:15 p.m., Lobby, Erickson

Transportation to the University Inn will be provided by IRT staff. Meet in front of Erickson at 5:15 p.m.

EVENING

Dinner - on your own.

From 8:30 p.m. - Invitation to: The Shulman's
1951 Penobscot
Okemos
Phone: 349-9469

Those in need of transportation may leave word with the IRT secretary in Room 133 Erickson.

FRIDAY, November 19

8:30 - 12 noon, 133 Erickson

Chairpersons: A. Porter and J. Stallings

Recorders: S. Olejnik and S. Wildfong

Discussion Topic: .OUTCOME MEASURES

Summary: N. L. Gage